

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:26:08 ; Search time 41.26 Seconds  
(without alignments)  
3391.977 Million cell updates/sec

Title: US-09-715-876-8  
Perfect score: 6495  
Sequence: 1 MLQQTLLFLYLFIASAKTI.....SIQHSTWLXLTLLSLFI 1260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

arched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneqseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	938	14.4	1537	15 AAR60562	Yeast 4.7 kb agglu
2	721.5	11.1	5179	22 AAM24516	C899P predicted am
3	685	10.5	2586	22 ABB66878	Drosophila melanog
4	640.5	9.9	2344	22 AAU37120	Staphylococcus aur
5	601	9.3	1795	22 ABB69806	Drosophila melanog
6	553.5	8.5	2870	21 AAY95559	Caenorhabditis ele
7	553.5	8.5	3178	21 AAY95556	Caenorhabditis ele
8	545.5	8.4	957	21 AAY59288	Human MUC11 polyt
9	545.5	8.4	957	22 AAM24513	C900P predicted am
10	534	8.2	849	17 AAR60725	F1Q1 protein, invo
11	516	7.9	894	15 AAR7578	Flocculation prote

12	516	7.9	894	15	AAR58754	S. cerevisiae FLO1
13	507	7.8	1532	21	AAB40945	Human ORFX ORF709
14	488	7.5	1721	19	AAB48299	Cryptosporidium pa
15	488	7.5	1721	21	AAB11727	Portion of Cryptos
16	487	7.5	1837	21	AAB11726	Cryptosporidium pa
17	468.5	7.2	862	15	AAB60563	Yeast 2.6 kb egglu
18	463	7.1	688	22	ABB30137	Peptide #2788 enco
19	463	7.1	688	22	ABB35307	Peptide #2813 enco
20	463	7.1	688	22	ABB20749	Protein #2748 enco
21	463	7.1	688	22	AAM56138	Human brain expres
22	463	7.1	688	22	AAM68511	Human bone marrow
23	463	7.1	688	22	AAM16315	Peptide #2749 enco
24	463	7.1	688	22	AAM28810	Peptide #2847 enco
25	463	7.1	688	22	AAM04053	Peptide #2735 enco
26	450.5	6.9	4498	22	ABF58595	Drosophila melanog
27	435.5	6.7	1045	22	ABG08332	Novel human diagno
28	428	6.6	1237	21	AAV81609	Streptococcus pneu
29	417.5	6.4	746	22	ABF59201	Drosophila melanog
30	417	6.4	560	22	AAU37464	Staphylococcus aur
31	404.5	6.2	2035	15	AAR57141	Host cell factor p
32	404	6.2	2570	22	ABG06375	Novel human diagno
33	402	6.2	2112	22	ABG60403	Drosophila melanog
34	392	6.0	502	22	AAU34408	Staphylococcus aur
35	392	6.0	1638	20	AAV00138	Enterococcus faeca
36	392	6.0	1638	20	AAV00142	Enterococcus faeca
37	392	6.0	596	21	AAV99408	Human PRO Polypt
38	390	6.0	596	22	AAU29178	Human PRO1342 (UNO
39	390	6.0	596	22	AAU29178	Human PRO1342 (UNO
40	390	6.0	596	22	AAU29178	Human PRO1342 (UNO
41	390	6.0	596	22	AAU29178	Human PRO1342 (UNO
42	379	5.8	1872	22	ABG15263	Protein of the inv
43	378.5	5.8	1976	22	ABF69419	Novel human diagno
44	377	5.8	2858	22	ABF71150	Drosophila melanog
45	377	5.8	3060	22	ABF58064	Drosophila melanog

ALIGNMENTS

RESULT 1  
ID AAR60562 standard; Protein: 1537 AA.  
XX AAR60562;  
AC AAR60562;  
DT 13-APR-1995 (first entry)  
XX  
DE Yeast 4.7 kb agglutination gene FLO1L.  
XX  
KW Yeast: agglutination; FLO1L.  
XX  
OS Saccharomyces cerevisiae ABXL-1D.  
XX  
XX WO9419475-A.  
XX  
PD 01-SEP-1994.  
XX  
PF 24-FEB-1994; 94WO-JP00290.  
XX  
PR 26-FEB-1993; 93JP-0038871.  
XX  
PA (PANI-) PANIMOLABORATORIO BRYGGERILABORATORIUM.  
XX (SABP) SAPPORO BREWERIES.  
XX  
PI Keraenen S, Ogawa M, Onnela M, Penttilla M, Takata Y;  
XX Watari J;  
XX  
DR WPI: 1994-294338/36.  
XX N-PSDB; AAQ71390.  
XX  
PT New yeast agglutination genes and yeast contg. them - impart  
PT agglutination properties to facilitate removal from fermentation  
media

xx  
PS  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
CC  
xx  
SQ

Disclosure; Page 43-48; 75pp; English.

The agglutination gene is called FLO1. Saccharomyces cerevisiae includes an agglutination gene of 4.7 kb (FLO1L) and an agglutination gene of 2.6 kb (FLO1S). FLO1L is the intact FLO1 gene on chromosome I, and FLO1S is the FLO1 gene with a portion of the ORF deleted in frame. FLO1L imparts a relatively strong CC agglutinative property to the host yeast into which it is introduced, while FLO1S imparts a weaker agglutinative property.

Sequence 1537 AA;

Query Match 14.4% Score 938; DB 15; Length 1537;  
Best Local Similarity 26.0% Pred. No. 3e-40;  
Matches 417; Conservative 244; Mismatches 523; Indels 418; Gaps 85;

QY 2 LQOFTLLFLYLSASAKY-----ITGVDFDSF-----NSLTWSNAANYAF-----40  
DB 10 lavftllal-tavagateacelpagqrkgmminfyqyslkdsstysnaaymaygysakt 68  
41 -----KPGVPTW-----NAVLGWSLD--G 58  
DB 69 klgsvggqtdisidynipcvsssgtfcpcqedsygnwckgmgaacnsqgiaywstdlfg 128  
QY 59 TSANPCDPTFLNN-----PCVFKVTTTQ-----TSVD 85  
DB 129 fytptnv-tlemtygflppqtgysytkfatvdasallsvggatafncccaqqqpilctn 187  
QY 86 LTADGVK-----VATCOFYSGEETFTSTLTCTVNDALKSSIKAFGT-----VTLP 131  
DB 188 ftldgikpvggalppniogtyvnygyyypm-----kvvyssnavsavgplpislvtlp 238  
QY 132 IAFNVGCTSSDLDKSCFTAGTNVFTVNDGDKDISIDVEFEKSTV-DPSAYLVASRVM 190  
DB 239 -----dgtvsvdfeg-----yyvsfdd-----disqncvtvpdsnya-vsttt 277  
QY 191 PSLNKKVTLFLVAPOECNYSCTGMFGSSNGGDAJDCSNTHIGITKGLNDWNPVSSSEF 250  
DB 278 titetpwtgtfst 337  
QY 251 SYTKTCTSGIOLKYNVPACYPFDDIAISATDVNOYTLAYNDVTCAGSRLOSKPFTL 310  
DB 338 elttvtgtngvtdetii-----vtrpttattaittepnwnstfststststststst 384  
QY 311 RWTGKNSDAGSNGI-----VIVATRTVTDSTTAVTTL-PFNPSVDKTKT-----355  
DB 385 -----ttvtgtnglptdetiivirpttattattttqpwndtstststststststst 438  
356 -----REILOPIPTTTT-----SVGVYTSYLTKTA-----PIGTATVIVDVPYH 398  
DB 439 ptdetliivirpttattattttqpwndtstststststststststststststststst 497  
QY 399 TTT--VTSEWGTIT-----TTTTRTN--PTSDIDTVVQVPLNPPT-VSTTEYWSQS 447  
DB 498 attamttqpwndtst 556  
QY 448 FATT--TTVTAPP-GTD--TVIIRPPNPT--VTTTEYWSQSFAAT--TTVTAPP 494  
DB 557 fst 616  
QY 495 G-GTDS--VIIRPPNPT--VTTTEYWSQSFAAT--TTVTAPP-GTDS--VIIRPP 542  
DB 617 glpdtetliivirpttattattttqpwndtstststststststststststststst 676  
QY 543 NPT--VTTTEYWSQSFAAT--TTVTAPP-GTDS--VIIRPPNPT--VTTTEYWSQS 591  
DB 677 tattamttqpwndtst 736  
QY 592 YATT--TTVTAPP-GTD--TVIIRPPNPT--VTTTEYWSQSFAAT--TTVT--635  
DB 737 fst 796

QY 636 GPSGTDTVLIREPNNP--TVTTEYWSQSFAAT--TTITAP--PCETDVTVLIREPP 686  
DB 797 gqptdetviivirptstseglvttttptgkgtfstststststststststststststst 856  
QY 687 NHTV--TTEYWSQSFAAT--TTVTAPP-GTD--TVIIRPPNPT--VTTTEYWSQS 735  
DB 857 seglittetpwtgtfst 916  
QY 736 YATT--TTVTAPP-GTD--TVIIRPPNPT--VTTTEYWSQSFAAT--TTVTAPP-GT 785  
DB 917 fst 976  
QY 786 -----DVIIVIESMSSSKI-----STSSNDITSIPSPSRP-----816  
DB 977 gqptdetviivirptstseglittetpwtgtfstststststststststststststst 1036  
QY 817 --HYVNSTTSD--LSTFSSSMNPTFSISSDGLMSSLTTLVTESETT--TELICSDGKE 869  
DB 1037 seglvttttptgkgtfst 1096  
QY 870 CSRLSSSGIVTNP--DSNESSIVTPTASTMDSLSSTGDISATSSDNNVSKGSVMT 927  
DB 1097 tssttsrpllt-pfypsnqtsvissvssvssstssstssstssstssstssstssst 1155  
QY 928 TETSVTITQTPNPLSSVTSLSLTOLSSIPSVSESKVFTFTNGDNQSGTHDSQSTSTET 987  
DB 1156 sessksv--lptssstsgssesetsagvs--sfisesakptyss--1 1205  
QY 988 EIVTTSST--KVLPPVSSNTDLTSEPTNTREOPTTLSTTS-----NSTEDITT 1035  
DB 1206 pvtst 1256  
QY 1036 SOPTGNDNTSSN--PVPTVATSTLASSEEDNKSGHESASTSLKPSMGNSGLTTS 1093  
DB 1257 tatvtvsgvteytwcpisttctckgkgtctctkqvtvtvsscesdvcsaktas 1316  
QY 1094 TETATTTSPTEAPSPAVSSGTDVTEPTDREOPTTLSTT-----SKTNSLVLAT-- 1145  
DB 1317 palvst 1374  
QY 1146 -----QATNENGKSPSTDLTSSLTGT--SASTSANSELVTSVGTGA 1188  
DB 1375 tatvndvvtvptwprptaneesvskmnsatgettntlaaettntvaetittntgaa 1434  
QY 1189 -----VASANDQSHSTV-----TNSNSIVSNTPOTTLSSQOVTS 1224  
DB 1435 etktvtstslsrshnaetqaaatdvighssvsvvsvaetgntckaitssgslstmaqqprst 1494  
QY 1225 SP-----STNTFIATYDQSGSIIOHSTWLYGLITLLSLFI 1260  
DB 1495 passmvgystaslelstysagansilaggslsvfiaillal 1536  
RESULT 2  
AAM24516  
ID AAM24516 standard; Protein; 5179 AA.  
AC AAM24516;  
XX  
DT 12-OCT-2001 (first entry)  
XX  
DE C899P predicted amino acid sequence.  
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
KW immunogenic; gene therapy; vaccine; colonic cancer.  
OS Homo sapiens.  
PN W0200149716-A2.  
XX  
PD 12-JUL-2001.  
XX



**RESULT** 3  
AB66878 standard; Protein; 2586 AA.

XX AC	AB66878;
XX DT	26-MAR-2002 (first entry)
XX DE	Drosophila melanogaster polypeptide SEQ ID NO 27426.
XX KW	Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
XX OS	Drosophila melanogaster.
XX WO	20001171042-A2.
XX DP	27-SEP-2001.
XX PF	23-MAR-2001; 2001WO-US09231.
XX PP	23-MAR-2000; 2000US-191637P. 11-JUL-2000; 2000US-061115O.
XX PA	(PEKE ) PE CORP NY.
PI	Venter JC, Adams M, LI PHD, Myers EW; WPI; 2001-656860/75. N-PSDB; ABL10981.
DR DR	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
PT PT	Disclosure; SEQ ID NO 27426; 2lpp + Sequence Listing; English.
XX CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL3051), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (AB857737-AB872072).
CC CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC CC	Sequence 2586 AA:
Query Match	10.5%; Score 685; DB 22; Length 2586;
Best Local Similarity	24.7%; Pred. No. 6.8e-27;
Matches 405; Conservative 183; Mismatches 557; Indels 494; Gaps 67;	
QY 15	ASAKTITTCVF--DFNSLTSWSNAANYAFKPGCYPTNWAVLGWSL--DGTSANPCDFTFLNM 71
Db 102	astntltkldddglsqqtcsasa-----pvvdvtggssngdgnstqsstttt--- 149
QY 72	PCVFKYTTSOTSVDLTADGVKYATCFQFVSGEETFTSTICTVNDAI---KSIKAFGTV 128
Db 150	-----tlttsstd-----ggefittssdpvvevsqgtngngnstqsgsst 188
QY 129	TUPTAFNVGGTGSDTD-----LEDSKCTAGTNTVPFNDGDKDISD-- 170
Db 189	tlttsdeggatssdpvvevaqgssngdgnstqslttttttsdggqstssdpv 248
QY 171	VFEKSTVDPSAYLYAS----RVMPSLNKVTTLFVAPOCE-----NG-----YTS 211
Db 249	vevsqgtngngnstqsgssttlttsdegqdtssdpvvevaqgssngdgnstqstt 308



CC for antibiotic development. The antisense nucleic acids can also be used  
CC to identify proteins used in proliferation, to express these proteins,  
CC and to obtain antibodies capable of binding to the expressed proteins.  
CC The proteins can be used to screen compounds in rational drug discovery  
CC programmes. The antisense nucleic acid sequences are also useful to screen  
CC for homologous nucleic acids which are required for cell proliferation in  
CC a wide variety of organisms. The present sequence represents an  
CC essential prokaryotic cellular proliferation protein.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ sequence 2344 AA;

Query Match 9.9%; Score 640.5; DB 22; Length 2344;  
Best Local Similarity 22.2%; Pred. No. 1.2e-24;  
Matches 323; Conservative 269; Mismatches 542; Indels 323; Gaps 46;

QY 17 AKTITGVDFNSLWNAAYAFKPGYPTWNAVLGWSLDGTSANPGDTFTLMPCEVK 76  
DB 297 ailtgridsnksfhgkvnlgnyeghngdgigaf-----spg----- 339

QY 77 YTTSTQSVDLTADGVKATCOFYSGEFTFTSTLTCTVNDALKSSIK-----AFGTV 128  
DB 340 -vigtglngaavgiglsnaf--gfkldtyhn--tstpnssakakadpsnvagggafgaf 395

QY 129 TLPIAFNVGSGTSDLED--SKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYA 186  
DB 396 vttdsygvastytsstsdnaaklnvgptnn--tfqd-----fdinyngdt----- 439

QY 187 SRVMPSLNKVTLFLVAPO-----CENGYTSGTMGFSSNGDVAIDCSNIHIGITK 236  
DB 440 -----kvmtkyvagqwtwnrnsdwlaksgtntfalsmtastg--atnlqqvqf-- 487

QY 237 GLNDWNPVSSSEFSYTKTCTSGIYKQYONVAG-----YRPFIDAYISATDVNOY 288  
DB 488 -----tfeytesavt--qrvydvvtgkdlipktyagndvqvtidnqqa 532

QY 289 TLATNDYTCAGSRLOS-----KPFTLRWT-----GYKNSDAGS-- 322  
DB 533 ltakgynytsavdsyastyndntkvmtnagqsvtyfcdvkaptvtvgntqievgtcm 592

QY 323 NGIVIVATRTVDTSTAVTTLTPNPSVDKTKTIBILOPI-----PT 364  
DB 593 npilvtltdngtgvtnvtvglpaglsydsatnslgtptkkgqstvtvatdqqnkt 652

QY 365 TTITTSVGVVTSYLTKTAPIGETATVI----- 392  
DB 653 tftlinvdtaptvt--pigdkssevfspislniatqdsngnavntvtgipsgltf 709

QY 393 -----VDVPYHTTTVTSEMTGTTTT-----TRTNPDTSDIDTVVQVPLPNP 436  
DB 710 dntnntsgtptnltstivtsdasgntktftkyevtrnmsdsvst--sgstqqq 767

QY 437 TVSTTEWWSQSFATT-----TTVTAPPGTDTVIIRE--PPNHTVTTEWWSQSFATT 488  
DB 768 svstskadsgaststsgsintstskstsvalsdsvsaaakslstseensvssstst 827

QY 489 TVTAPPGTDSVIIREPPNPVTVTTEWWSQSFATTTVTAPPGGTDSVIIREPPNPVTVT 548  
DB 828 slvnsqsvss--mgsvsakstslsdflnsstskstseststslstslstsdsvm 885

QY 549 TEWWSQSVATTTVTAPPGGTDSVIIREPPNPHTVTTEWWSQSVATTTVTAPP----- 601  
DB 886 stsgslsksgslstsdsgstsdsgstsdsgstsdsgstsdsgstsdsgstsdsgst 945

QY 602 -----PGTDTVIIREPPNPHTVTTEWWSQSFATTTVTAPPGGTDSVIIREPPNPVT 654  
DB 946 sastsklesgslstsdsgstsdsgstsdsgstsdsgstsdsgstsdsgstsdsgst 1005

QY 655 TTTEWWSQSVATTTVTAPPGGTDTVIIREPPNPHTVTTEWWSQSVATTT-----TTVT 707

Db 1006 stsem1sdsmstsglaaadsksmsvssmstsgstseslsdsstsdskdsst 1065  
QY 708 APPGTDVTL-----IREPNHTVTTEYWSQSYATTTVTAPPGTDTVIIRPNPTV 762  
Db 1066 sqsgstststassvmaesqsgstsmstsgstsdstsdstsdstsdstsdst 1119  
QY 763 TTTEYWSQSFATTTVTAPPGTDTVIIRPNHTVTTEYWSQSYATTTVTAPPGTDTVIIRPNPTV 822  
Db 1120 asseisqsgststsdstsdstsdstsdstsdstsdstsdstsdstsdstsdst 1174  
QY 823 TSDLTSTFSSMNTPTGIS-----SDGMLLSSTTLVTE-----SEMTTELCSDGKE 869  
Db 1175 tad-aiseisgaesteiseseenstsdseekaafiseiseseestseeslsgstsd 1233  
QY 870 CSRLS-----SSGIVNPDSNDSSTVTPTAST-----MPSLSLS 907  
Db 1234 stslsdnsesgststslsdnsesgststslsdnsesgststslsdnsesgststslsd 1293  
QY 908 TDCISATSSDNVKS-----GVSVTETSTVTIQTTPNPLSSSVTLTQLSSIPSVSESE 962  
Db 1294 stslsdstsdsgsknsalsasmstsdstsksealsaatstslsdstsdsgstse 1353  
QY 963 SKVTFTS-----NGDNQSGTHD--SOSTSTETETETVTSTTKVLPVVSNTDLTSE 1011  
Db 1354 sksdstsmislsqsgstsgstsvstsealsdststslsdstslsdstslsdstslsd 1413  
QY 1012 PNTRE--QPTTLSTSTNSITEDITTQPT--GDNDNTSSTNPVPTVATSTLASSEED 1067  
Db 1414 stslsdsgstsgstsgstsgstsgstsgstsgstsgstsgstsgstsgstsgstsgst 1473  
QY 1068 NKSGS-----HESASTSLKPSMG--ENSGUTTTETETATTTSPTEAPSPAVSSGTDV 1117  
Db 1474 sdsgststslsdsgststslsdstslsdstslsdstslsdstslsdstslsdstsl 1532  
QY 1118 TTEPTDREQPTTLSTKTNSELVAT-----TQATNKGKSPSTDLTSSLT 1165  
Db 1533 lsasmsesdsqstslsdstslsdstslsdstslsdstslsdstslsdstslsdstsl 1592  
QY 1166 TCTSASTSANSSELVSGSVTGAVASANDQSHSTVNSNSIVNTPORTLSQOVTSSS 1225  
Db 1593 tslsdsgstsdstsdsgststslsdstsdstsdstsdstsdstsdstsdstsdst 1644  
QY 1226 PSTNTFIATYDGGSGSI 1242  
Db 1645 tststsvdsstslsdsv 1661

RESULT 5  
ABB69806 standard; Protein; 1795 AA.

AC ABB69806;  
XX 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 36210.  
DE Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX Drosophila melanogaster.  
OS  
PN WO200171042-A2.  
PD 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
XX 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI; 2001-65860/75.  
DR N-PSDB; ABLI3909.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions -  
XX Disclosure; SEQ ID NO 36210; 21pp + Sequence Listing; English.  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA  
XX sequences (ABLI01840-ABLI16175) and the encoded proteins  
XX (ABBS7737-ABBS70272).  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 1795 AA;  
SQ  
Query Match 9.3%; Score 601; DB 22; Length 1795;  
Best Local Similarity 24.7%; Pred. No. 9.4e-23;  
Matches 323; Conservative 162; Mismatches 499; Indels 322; Gaps 57;  
QY 126 GTVTLPIAFNV--GGTSGSTDLSDSKCTAGTNTVTNFGDKDIDIVFEKSTVPSAY 183  
Db 182 glifsvrklpdcpcstelsdsgyqncelkfpceaeegtf-----rptdcaly 236  
QY 184 LYASVMPNLKVTTLFVAPQENGVTSGTWGSSNGDVAIDCSNIHIGITKGLNDWNY 243  
Db 237 -ytrfiesgtylqtrfkcp----gsnfdlerklorprsevdcdcfvpg----- 281  
QY 244 PYSSSESRYKTCSTNGIOIKYQNPAGYRPFDAISATDVNQVTLAYTNDYTCAGSR- 302  
Db 282 pv-----qvyepqpyppypaappl-----yeeddydgcare 314  
QY 303 ----LOSKPFTLRWTKYKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVD----- 351  
Db 315 qpalksekliqaaegfep---slnvvviqt--tleptayhkypayspsysesh 369  
QY 352 -----KTKIE--ILQP-IPTTTTTT-----SYVGVTTSYLT 380  
Db 370 hrkeraaenlekegvrklkseniqlpdpataatprelndlnkygkrytygt 429  
QY 381 KTAIGETATV-----IVDPVYHTTTVTSEWGTGITTTTTT---TNPTDSID 425  
Db 430 dkndvteapeikspklhlsenivilpettttttttkpvlcptisppdttkpstk 489  
QY 426 TVVQVPLPNTVSTTEYWSQSFATTTVTAPPGTD--TVIIRPNHTVTTEYWSQSF 484  
Db 490 cavtk---scpklsstceqstttakttkrtvttektssatekprttvtvttt--tqkr 544  
QY 485 ATTTVTAPPGGDSVIIRPNPTVTTEYWSQSFATTTVTAPPGTDSVIIRPNP 544  
Db 545 sttChmisp--dtktlrirstlspktttpt-----sttstpttptst-----tp 587  
QY 545 TVTTEYWSQSFATTTVTAPPGTDSVIIRPNHTVTTEYWSQSFATTT--TVTAPP 602  
Db 588 sttpt-----sttstpttptsttttkv---sthrpttcsqkttaasttkkttkgpk 637  
QY 603 GGTDTVIIRPNHTVTTE---YWSQSFATTTVTGPPSGCTDTVIIRPNPTVTTEY 659  
Db 638 tkttdlptstskstslsttkttttkhkaattsttkpstk-----tektatvatkttk 692  
QY 660 WSQSYATTTTAPPGTDTVIIRPNHT--VTTTEYWSQSFATTTVTAPPGTDT 714

Db 693 atesspktsstgkpttkpkpstrttpttkvttttq1tttltplrsstettstqtpptt- 751  
 QY 715 TVLIREPPNHTVTTTEYWSQSYATTATVTPGCTDVIIIREPPNHTVTTTEYWSQSPAT 774  
 DR 752 -----tpqptttltvtpktsttttttkp-----itaspkpt-tttqtktttaepn 797  
 QY 775 TTTV-----TAPGCTDVIIRESKSSKISSTND--TSLIIPSPSPHYVNSFTSD 825  
 PT 798 tkkvaltkettpkgstttftrktttnpepistekpistcp---kp---sttkp 851  
 QY 826 LSTPSSSMNTPTSSSDGMLLSTVLTETSEPTTELICSDKECSRSSSSGIVNPDPS 885  
 Db 852 Lstvasst--ekttisap---kptekstentpnsv-----kcalststqratstts 900  
 QY 886 NESSI---VTSVTPASTMSDLSLSDIGSATSSDNYSKSGVSVTE-----TSVVTI-- 935  
 Db 901 epttktnlttkpklitkstqe---atstqkvst--vlttkkatespklitst 954  
 QY 936 ---OTTPNPL-----SSVSTSLTOLSIIPSVSESSEKVTFTSNGDNGTHDSQST--- 983  
 Db 955 eepntkplrtrttttsvtatritt--tlisestctstqkpkstcpsttrtkpv 1013  
 QY 984 -----STEIEIVTSSKVLPPVSSNDLSEPTNTRQPT--TLSTTSNSTRIDITTS 1036  
 Db 1014 ttvlstqpttttkstvtlttpppsqqrpttcttrqptsicastsigttripttt 1073  
 QY 1037 QP-----TGCGDNSTNPNVPFVATSTLASAEEDNKGSGHE 1074  
 Db 1074 npqntssldtlvtrppcpdpdstskntactqelqgv--nillelqspkqeqfht 1131  
 QY 1075 SASFSLKPSMGNSG-----LTSFEIATVTSPTSPAVSSGTD-----V 1117  
 Db 1132 rthaltgarnltgqevpdydpmdapssaeasgqaktakaptmtslaaahllqklfhl 1191  
 QY 1118 TTEPTDTRQPTTLSTTSKTNSEL-----VATTO-----ATNE 1150  
 Db 1192 sttpssreahqrpssqgassgrgtiagmarhnlatskpfiahslrlslqqlastq 1251  
 QY 1151 NGKSPSTOLTSSTLTGTSASTSANSSELVTSVTCGAVASANDQSHSTSVNSNSIVS 1210  
 Db 1252 krsipkktlvthntckepedseydydse--tsegytdednevladkqpramssttvaavlp 1309  
 QY 1211 NTPQTLISO--QVTSSTSPNTPIAST-----YDGSGS 1241  
 Db 1310 avpattterepktsasepsptkatsatttqptlettgdldleydsags 1355  
 RESULT 6  
 ID AAY95559 standard; Protein: 2870 AA.  
 AC AAY95559;  
 XX 10-OCT-2000 (first entry)  
 XX Caenorhabditis elegans LOV-1 deletion allele lov-1(sy582del).  
 DE Location of vulva; LOV-1 gene; nematode; mating behaviour;  
 KW polycystin; polycystic kidney disease; animal model;  
 KW signal transduction; mutant; mutain.  
 XX Caenorhabditis elegans.  
 OS Synthetic.  
 XX WO2000040711-A2.  
 PN 13-JUL-2000.  
 PD 06-JAN-2000; 2000WO-US00521.  
 PR 06-JAN-1999; 99US-0115127.  
 XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.

XX Sternberg PW, Barr MM;  
 PI WPI: 2000-452537/39.  
 XX Isolated nucleic acid molecules from Caenorhabditis elegans useful for  
 XX producing transgenic nematodes with altered mating behavior for  
 XX identifying genes or regulatory factors involved in polycystic kidney  
 XX disease -  
 XX Example 1; Page 131-139; 142pp; English.  
 XX The present sequence is that of a deletion allele, termed  
 CC lov-1(sy582del), of the Caenorhabditis elegans LOV-1 protein  
 CC (see AAY95556), generated by genomic deletion of the PKD/channel  
 CC domain of the lov-1 gene by PCR. LOV-1 is an orthologue of  
 CC the human polycystin-1 (PKD1) gene that is defective in human  
 CC autosomal dominant polycystic kidney disease (PKD). The LOV-1  
 CC protein is involved in chemosensory or mechanosensory signal  
 CC transduction in sensory neurons. It is required for 2 male  
 CC sensory behaviours, 'response' of males to hermaphrodites, and  
 CC 'location' of the vulva. Nematodes such as C. elegans that  
 CC express mutant or wild-type LOV-1 or PKD-2 (see AAY95557) can be used  
 CC to study the functions of the proteins encoded by these genes, to  
 CC screen for other genes involved in PKD, to identify mutations  
 CC involved in the disease, and to screen for drugs that affect PKD.  
 CC Behaviours controlled by the action of the genes or gene products  
 CC are identified and used in the assays. Hence, an animal model is  
 CC provided that permits study of the etiology of PKD and provides a  
 CC tool to identify the genes involved in the disease pathway, to  
 CC identify compounds that may be used to treat or alter the disease  
 CC progression, lessen its severity or ameliorate symptoms.  
 XX Sequence 2870 AA;  
 SQ  
 Query Match 8.5%; Score 553.5; DB 21; Length 2870;  
 Best Local Similarity 24.3%; Pred. No. 4.8e-20;  
 Matches 290; Conservative 154; Mismatches 401; Indels 349; Gaps 51;  
 QY 252 YTKTCTSGNIQIKYQNPAGYRPFIDAVISATDVNQYTLAYTNDYTCAGSLQSKPFTLR 311  
 Db 194 yrekcesgeineeyarimc-kryprse--kstaidsdsgqvydgqvlkgr--akqfsmr 248  
 QY 312 WTG-----YKNSDAGSNGIVIVATRTVTDSTTAVTTLPNPSPVDKTKTEILOPIPTT 366  
 Db 249 tsdgsptlrmkrdagdn-----tcdytilestststttptttvtstvt-----sttt 295  
 QY 367 ITTSYGVVTSYLTAKTAPIGETATVIVDVPVHTTTVTSEMTGTTTTRTNPTSDIT 426  
 Db 296 vptststvtamstststststt-----eststfts-----taststststtqass 346  
 QY 427 VVQVPLPNPTVSTTEYWSQSFATTT--VTAPPGCTDVIIIREPPNHTVTTTEYWSQSF 484  
 Db 347 titssp-sstlist-----spttttpeiststlsslpdnalcyldeittstfttml 399  
 QY 485 ATTTTVTAPPGCTDVIIIREPPNHTVTTTEYWSQSFATTTVTAPPGCTDVIIIREPPNP 544  
 Db 400 ttttt-----eepststttevtststvttt-----epttt 431  
 QY 545 TVTTEYWSQSYATTTVTAPPGCTDVIIIREPPNHTVTTTEYWSQSYATTTVTAPPGG 604  
 Db 432 lltst 475  
 QY 605 TDVVIIREPPNHTVTTTEYWSQSFATTTVTGP-----PSCDTVVIIREPPNHTVTTTE 658  
 Db 476 tscest 530  
 QY 659 YWSQSYATTT--TITAPPGCTDVIIIREPPNHTVTTTEYWSQSYATTTVTAPPGG 702  
 Db 531 --tqgssttkststsdgtdnqdfyve---katttfydstsvnltnslgllgvt 584  
 QY 703 TTTVTAPPGCTDVIIIRE-----PPNHTVTTT----- 729





QY	823	TSDLSTFESSNMNTPSTSSDGMLLSSSTLTVTSEPTTELICSDKCEKSLSSSSGIVTN	882
Db	492	sta fcthpasttttptst-----pstatapveestcyhrspstctptthfpass---tt	540
QY	883	PDSSNESSIVTSPNPAAS--TWSDSLSSSDGISATSSDNVSKSGVSVTSTVTTQTTPN	940
Db	541	aghaekstifhspdagstcpssahattscgreettsrjspstetltlpgst---tlpg	597
QY	941	PLSSSVTSTQLSSSTPSSSESKVTFVTSNGDNOSGTHDSQSTSTETETIVTSSKVLPP	1000
Db	598	-lseastfysprsppttlispasmtslgvge-esttsrsgpgsthatvcpastt---tp	652
QY	1001	VYS--SMTDLTSEPNNR-----EQPTTL-----STGNSITEDITTS---	1036
Db	653	glseestfvysspgstctctvfratttsvrgceptfhrpashttlftedstsglt	712
QY	1037	-QPTGDNQNTSTNPVTVATSTLASASEDNK-SCSHESASTSLKPSMGENSEGL---	1090
Db	713	eestafpgspastqglp--attltadlgeestfhpsssgsgtltsparstsglvges	770
QY	1091	-----TTSTETETATTT-----SPTETAPSPAVSGVDVTTPTPTDTR	1125
Db	771	tpsrlepstettlpgspcttpsekatftfysrpdattlspatttsgvseestsh	830
QY	1126	EQPTTLSTGTS-----KTNSLTVATTQATNNGCKSPDTLTSSL-----	1164
Db	831	sqpgsthtcafpdcttsglsgpektshssgsgtcat-----lspgstatsslgqgsttf	885
QY	1165	--TTGTSASTNSANSELVTSQVTVGGAVASANDOSHSSTSVTNSNSIVSNPTQLSQOVT	1222
Db	886	hspsgpdtettlppdtstslgveastphtsatsglhttltpasttsaglgqeestfqsqp	945
QY	1223	SSSPST	1228
Db	946	sssdtt	951
RESULT	9		
AAIM24513			
ID	AAIM24513	standard; Protein; 957 AA.	
AC	AAIM24513;		
DT	12-OCT-2001	(first entry)	
DE	C900P	predicted amino acid sequence.	
KW	Human; immunotherapy; diagnosis; colon cancer; colon tumour;		
	immunogenic; gene therapy; vaccine; colonic cancer.		
	Homo sapiens.		
XX	WO200149716-A2.		
PN	12-JUL-2001.		
PD	29-DEC-2000; 2000WO-US35596.		
XX	30-DEC-1999; 99US-0476296.		
PR	10-JAN-2000; 2000US-0480321.		
PR	15-FEB-2000; 2000US-0504629.		
PR	06-MAR-2000; 2000US-0519444.		
PR	19-MAY-2000; 2000US-0575251.		
PR	29-JUN-2000; 2000US-0609448.		
PR	28-AUG-2000; 2000US-0649811.		
XX	(CORI-) CORIXA CORP.		
XX	Xu J, Lodes MJ, Secretist H, Benson DR, Meagher MJ, Stolk JA;		
PI	King GE, Wang T, Jiang Y;		
XX	WPI: 2001-441847/47.		
DR			

Colon tumor associated proteins and nucleic acids useful for the prevention, diagnosis and treatment of colonic cancer -

Claim 2: Page 437-440; 472pp: English.

The present invention describes colon tumour associated proteins (I) and the polynucleotides (II) that encode them. (I) have cytostatic activity. (I) and (II) can be used in gene therapy and vaccine production. (I) and (II) may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate colon tumour associated protein (TCAP) expression, such as colonic cancer. For example, (I) and (II) may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of TCAPs by expressing inactive proteins or to supplement the patients own production of them. Additionally, (II) may be used to produce the TCAP proteins, by inserting the nucleic acids into a host cell culturing the cell to express the protein. (II) and its complementary sequences may also be used as DNA probes in diagnostic polymerase chain reaction (PCR) and hybridisation assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (I) may also be used as antigens in the production of antibodies against TCAPs and in assays to identify modulators of TCAP expression and activity. Anti-(I) antibodies and antagonists may also be used to down regulate TCAP expression and activity. The anti-(I) antibodies may also be used as diagnostic agents for detecting the presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512 and AAM24494 to AAM24523 represent nucleotide and amino acid sequences given in the exemplification of the present invention.

Sequence 957 AA:

Query Match	8.4%	Score 545.5;	DB 22;	Length 957;
Best Local Similarity	25.9%;	Pred. No. 3.2e-20;		
Matches 266;	Conservative 139;	Mismatches 406;	Indels 215;	Gaps 43;

QY	338	TTAVTTLTFNPNPSVDKTKTIELLIQPIPTTTTTSYGVQVTTSYLTATGRTATVIVDVPY	397
Db	6	ttafpgsttmppgvsqestas---hspsgdsttllspgstta-----salspgesttffhsppg	58
QY	398	HTTTTIVSEWTT---GPIITTTTRNPNRDSIDTVVQVQPLNPRTVSTT-----EYWSQ	446
Db	59	stettllpndntasglleaestpvnstgsphttl-----spagsttqrqgesttffgwpn	112
QY	447	SFATTTVTVAPPGGTDTVIIREPPNH---TVTTTEYWSQSFAT-----TTVTVAPPGGT	497
Db	113	skdtt---papptttsafvelstshspsbstptthfaasattlgraeestttfhspsrvat	169
QY	498	DSVIIREPPNPVTITTEYWSQSFATTTTVPAPGGDSVIIREPPNPVTITTEYWSQSYA	557
Db	170	att-----psparstt---sglveesttyhsppgstqtmhfpesatctsgsgeesttshs	221
QY	558	TTTTVTAPPGGTDSVIIREP-----PNHTVTT-----TEYWSQSYA-----TTTT	597
Db	222	thtlhsap-sttsalveeptsyhsppgstcatthfpesttsgseestashangdatgt	280
QY	598	VTAPPGGTDTVIIREPPNHTVTTEYWSQSFATTT---TTVTGPPSGTDTVIIREPPNP	653
Db	281	lvlparsatttllvges-----ttspisgsmettalpgsttptglsokstfhspsrpa	335
QY	654	VTTEYWSQSYAT---TTTTIAPGEGTDTVIIREPPNHTVTTEYWSQSYATTTVTAPP	710
Db	336	ttlspasttsgsgeesttsrpggstthtcafops-----tttptglshr---sttshasp	387
QY	711	GETDTVIIREPPNHTVTTEYWSQSYATTTTVPAPGGTDTVIIREPPNPVTITTEYWSQ	770
Db	388	gstdtll---pas-----tttsgpsqe---sttshsppgsttatal-----spgsrsta	434
QY	771	SFATTTVTVAPPGGTDTVIIRESSSSKI-----STSSNDMTSIPFSRPHVNST	822
Db	435	e---sttffhsppgstthtllfpgdsttsglveastvrvhstsgpttctllpasskspglge	491



QY 1165 TTGTSASTSANSSELTSGSVTGAVASANDQSHSTVNSISVNTPTQTLSSQOVTS 1224  
 Db 756 srfnhaetqasatdvgh--sssvsvsetgntkglitsetlsmssqprstpastmvg 813  
 QY 1225 SPSTNTIASTYDGSISIQHSTWLYGLTL--LSLFI 1260  
 Db 814 --stasleistyvgian-----gillnngisvfi 840

RESULT 11  
 AAR47578  
 ID AAR47578 standard; Protein; 894 AA.  
 XX  
 AC AAR47578;  
 XX  
 DT 19-JUL-1994 (first entry)  
 XX  
 DE Flocculation protein of *Saccharomyces cerevisiae*.

DE Major cell wall protein; glycosyl-phosphatidyl-inositol;  
 KW anchoring protein; alpha factor; alpha-agglutinin; invertase;  
 KW inulinase; alpha-amylase; *Saccharomyces cerevisiae*;  
 KW flocculation protein; enzymatic process; fermentation;  
 KW biodegradation; catalysis.  
 XX  
 OS *Saccharomyces cerevisiae*.  
 XX  
 PN WO9401567-A.  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 07-JUL-1993; 93WO-EP01763.  
 XX  
 PR 08-JUL-1992; 92EP-0202080.  
 PR 14-DEC-1992; 92EP-0203899.  
 XX  
 PA (UNIL ) UNILEVER NV.  
 PA (UNIL ) UNILEVER PLC.  
 XX  
 PI Kllis FM, Schreuder MP, Toshchka H, Verrrips CT;  
 DR WPI; 1994-035071/04.  
 DR N-PSDB; AQ034029.

XX Immobilisation of enzymes to microbial cell wall - by prodn. of  
 PT fusion protein of enzyme linked to anchoring protein  
 XX  
 PS Example 10; Page 59-64; 99pp; English.

XX The flocculation protein is used in a method to immobilise enzymes  
 CC to a microbial cell wall. The coding sequence (FLO 1) is used in  
 CC the production of a recombinant polynucleotide which comprises a  
 CC structural gene encoding a protein with catalytic activity and at  
 CC least part of a gene encoding at least the C-terminus of a protein  
 CC capable of anchoring in a eukaryotic or prokaryotic cell wall. The  
 CC anchoring fragment or protein is selected from alpha agglutinin,  
 CC AGA 1, FLO 1, major cell wall protein of lower eukaryotes or a  
 CC proteinase of lactic acid bacteria. The recombinant polynucleotide  
 CC preferably also comprises a sequence encoding a signal peptide to  
 CC ensure secretion of the expressed product. The signal peptide is  
 CC preferably derived from glycosyl-phosphatidyl-inositol, anchoring  
 CC protein, alpha factor, alpha-agglutinin, invertase or inulinase,  
 CC alpha-amylase of *Bacillus* or proteinases of lactic acid bacteria.  
 CC The host microorganism can be used for performing enzymatic  
 CC processes on an industrial scale.  
 XX  
 XX Sequence 894 AA;

Query Match 7.9%; Score 516; DB 15; Length 894;  
 Best Local Similarity 26.3%; Pred. No. 9.8e-19;  
 Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

QY 437 TVSTTEYWSQSFATTTTAPPGGDTVIIREPPNHTVT-----477  
 Db 53 tysnaaymaygyaskklsgvggqtdisidynlpcvsssgtfcpcqedsygnwckgmga 112  
 QY 478 -----EYWSQS-FATTTTAVTAPPGTDSVIIREPPNPTVTTEY-----WSQSEA 521  
 Db 113 cansgglaywstldlfgfyt-----ptnvtlemtygfllpqtgtytkfa 157  
 QY 522 TT-TTAVTAPPGTDSVIIREPPNPTVTTEYWSQSFATTTTAPPGGTDSVIIREPPNH 580  
 Db 158 tvddsailsvggatafnccagqppitstnf-----tldgikpwggs-----lppnl 204  
 QY 581 TVTTEYWSQSY-----ATTTAVTAPPGT-----DTVI 609  
 Db 205 egtvymagyyvpmkvvynavswglplslvtlpgdtvtsdddfegyyvsfddlsqsnct 264  
 QY 610 IREPPNH---TVTTEYWSQSFATTTTATGP-----PSGTDTVIIREPPNPTV--TT 656  
 Db 265 vdpnsyavsttttttptwtgftststetmtvtgngvptdetvirltptaeqlstt 324  
 QY 657 TEYWSQSVATT-----TTITAPPE-TD-TVLIREPPNHTV--TTTEYWSQSVATT---- 703  
 Db 325 tptwtgftststetvttitgngqptdetvirltptseglsttptwtgftststem 384  
 QY 704 TTVTAPPGTDTVIIREPPNHTVTTEYWSQSVATTTTAPPGGTDTVIIREPPNP--T 761  
 Db 385 tltvtgng-----qpidelv-----ivrltptseglv 411  
 QY 762 VTTTEYWSQSFATTTTAPPGGT-----DTVIIYE---SMSSSKISTSSNDITSIP 811  
 Db 412 ttttptwtgftststetmtvtgnglptdetvirltptttdaislsssssgqitsslt 471  
 QY 812 SFSR----PHYVNSTTDLSTFESSMNTPTSISSDGLLSSTTLVTESETTELICSDG 867  
 Db 472 S-srplltcpfypngtvisssvisssvtlsfscapvi--ssvissttsttsifses 528  
 QY 868 KCSRPLSSSGIVTNPDSNESSIVSTVPTASTMSQSLSTGDIATSSDNVSKGVSVT 927  
 Db 529 sksviptsstsgssesetss-----agvsassfssesksptyssasip 576  
 QY 928 TETSVTTIQTTPNPLSSSVTLTQLSSIPSVSESEKVTFTSNGDNQSGTHDSQSTET 987  
 Db 577 lvtststqetasslppattkt-----seqtlvtv-----tscas 613  
 QY 988 EIVTTSSTKVLPPVSVSNTDLTSEPTNTREOPTTLSTTSNSITEDITTSQPTGDCDNYS 1047  
 Db 614 hvctes---lspalvatvrvsgvtteytwcpistce-----ttkqkgt-edlt 662  
 QY 1048 STNPVPTVANSTLASSEED--NKSGSHESASTSLKPSMGENSEGLTT-----STEIET 1100  
 Db 663 etkqktvtv---lsscesdvcskaspalvstcatngvtteytwcpistcsrqqt 719  
 QY 1101 TSPT-----EAPSPA-VSSGTD-----VTTEPTDTREOPTTLSTTSKTNSELVAT 1144  
 Db 720 tlvtvtscesgvcsetaspalvstatacvndvvtvptwprqtaneesvsskmsatget 779  
 QY 1145 TQATNENGKSPSTDLTSLTGTGTSAITSANSELVTSGSVTGG-----AVASASNOOSH 1199  
 Db 780 T--tntlaaettntvaaetintga--aetktvvtalsrshnaetqtasatdvighs 834  
 QY 1200 TSN--TNSNIVSTPQTTLSSQOVTSSTSP-----STNTFTIASTYDGS 1241  
 Db 835 ssvsvsvsetgntkltssglstmsqprstpsasmvgyastaslststygast 887  
 RESULT 12  
 AAR58754  
 ID AAR58754 standard; Protein; 894 AA.  
 XX  
 AC AAR58754;  
 XX  
 DT 27-MAR-1995 (first entry)



XX DE S. cerevisiae FLO1.  
 XX KW Binding protein; immobilization; chimeric protein;  
 KW anchoring protein; Saccharomyces cerevisiae; flocculation;  
 KW FLO1 gene, scfv; single chain antibody; monoclonal antibody;  
 KW MAB; human chorionic gonadotropin; HCG.  
 XX OS Saccharomyces cerevisiae.  
 XX PN WO9418330-A.  
 XX PD 18-AUG-1994.  
 XX PF 10-FEB-1994; 94WO-EP00427.  
 XX PR 10-FEB-1993; 93EP-0200350.  
 XX (UNIL ) UNILEVER NV.  
 XX (UNIL ) UNILEVER PLC.  
 PI De Geus P, Frenken LGJ, Kils FM, Toschka HY, Verrips CT;  
 DR WPI; 1994-279751/34.  
 DR N-PSDB; AAQ67360.  
 XX Immobilised binding proteins for specific cpds - obtd. by  
 PT expressing chimeric proteins comprising the binding protein and a  
 PT cell wall-anchoring protein in host cells  
 XX Disclousure; Page 36-39; 78pp; English.  
 XX Valuable compounds are isolated from complex mixtures by use of  
 CC immobilised ligands composed of an anchoring protein and a binding  
 CC protein. A suitable anchoring protein is yeast FLO1 (associated  
 CC with flocculation), and a gene encoding a chimeric scfv-FLO1 protein  
 CC that will anchor in the cell wall of a lower eukaryote, and which  
 CC binds HCG with high specificity, was produced.  
 XX Sequence 894 AA;

Query Match 7.98; Score 516; DB 15; Length 894;  
 Best Local Similarity 26.38; Pred. No. 9.8e-19;

Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

QY 437 TVSTTEWWSQSFATTTTVPAPGCTGTVIIRPPNPHVTVT-----DTVI 477  
 DB 53 cysnaaymgygasklgsvggqtdisidynlpcvsssgtfcpcqedsyngwckgmg 112  
 QY 478 -----EWSQS-FATTTTVPAPGCTDSVIIRPPNPHVTVTTEY-----WSQSPA 521  
 DB 113 csnqglaywscdlfgfyt-----pcnvtlemtygfllpqqtsgytkfa 157  
 QY 522 TT-TTVPAPGCTDSVIIRPPNPHVTTEWWSQSFATTTTVPAPGCTDSVIIRPPNH 580  
 DB 158 lvdslavagatafcacqaqppltstnf-----tidgikpwggs-----lppni 204  
 QY 581 TVTTEWWSQSFATTTTVPAPGCTGTVIIRPPNPHVTVT-----DTVI 609  
 DB 205 egtvmyagyygmkvvyvnavawgtlplsvtlpdgttvsddffegvyysfddlsqsnct 264  
 QY 610 IREPPNH-----TVTTEWWSQSFATTTTVPAPGCTDSVIIRPPNPHVTVT-----TT 656  
 DB 265 vpdpsnyavsttttctepwgttstststststststststststststststststst 324  
 QY 657 TEWWSQSFATTTTTPAPGCTDSVIIRPPNPHVTTEWWSQSFATTTTTPAPGCTDSVIIRPPNH 703  
 DB 325 lcpwtgftstevttitngngqtdetvirltptseglisttptgftststem 384  
 QY 704 TTVTAPGCTDSVIIRPPNPHVTTEWWSQSFATTTTVPAPGCTDSVIIRPPNP--T 761  
 DB 385 tvtgting-----qtdetv-----ivrltptseglv 411

QY 762 VTTEWWSQSFATTTTVPAPGCTGTVIIRPPNPHVTVT-----DTVIYE-----SNSSSKISTSSNDITSIP 811  
 DB 412 tttcpewtgftst 471  
 QY 812 SFSR---PHYVNSTDLSTFESSMMPTSISSDGMLLSLTTLVTESETTELLICSDG 867  
 DB 472 s-srplltfypngtvisssvltssalfstspvi--ssvltssstttststststst 528  
 QY 868 KECRLSSSSGIVTPDSSNESSIVTVPATSMDSLSLSTGIGISATSSDNVSKSGSVT 927  
 DB 529 sksviptststsgssesetss-----agvassafissesskptysaslp 576  
 QY 928 TETSVTTIQTTPNPLSSSVTLTQLSSIPSVSESKVTFSTNGDNQSGTHDSQSTSTEI 987  
 DB 577 lvtstststetasslppatttk-----seqttltvt-----tscs 613  
 QY 988 EIVTSTKVLPPVSSNTDLTSEPNTNREQPTTLSTTNSITETDITTSOPTGNDGNTS 1047  
 DB 614 hvctes---ispavstvtvsgvtteytwcpistte-----ttkqkgtt-eqtt 662  
 QY 1048 STNPVPTVATSTLASASEEP--NKSGSHESATSLKPMGENSEGLTT-----STEIEATT 1100  
 DB 663 ettqktvtv---lascsedvcaktaapaivstataingvteytwcpisttearqqt 719  
 QY 1101 TSPT-----EAPSPA-VSSGTD-----VTTEPTDTRQPTTLSTTSTKTNSELVAT 1144  
 DB 720 tlvtvtscsgyvcsetaspaivstataingvteytwcpisttearqqt 779  
 QY 1145 TQATNENGCKSPSTDLTSLTGTSTASANSSELVTSQVTCG-----AVASASNDQSHS 1199  
 DB 780 t--tntlaaettintvaetintga---aectvtvtslsrsnhaetqatstadvigns 834  
 QY 1200 TSV-----TNSNSIVSNTPPTTLTSLQVTVSSSP-----STNTFIATSYDGS 1241  
 DB 835 ssvvsvsetgntktsitsgistmsqqrstpsasmvgystaslelstysagat 887

RESULT 13

AA840945  
 ID AA840945 standard; Protein; 1532 AA.

XX AC AA840945;

XX DT 08-FEB-2001 (first entry)

XX DE Human ORF709 polypeptide sequence SEQ ID NO:1418.

XX KW Human; open reading frame; ORF; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.

XX OS Homo sapiens.

XX PN WO2000058473-A2.

XX PD 05-OCT-2000.

XX PF 31-MAR-2000; 2000WO-US08621.

XX PR 31-MAR-1999; 99US-0127607.

XX PR 02-APR-1999; 99US-0127636.





ID AAB11727 standard; Protein; 1721 AA.  
 AC AAB11727;  
 XX  
 XX  
 XX 28-OCT-2000 (first entry)  
 XX  
 XX  
 XX Portion of Cryptosporidium parvum NINC isolate GP900.  
 XX  
 XX GP900; NINC isolate; glycoprotein; antibody; cryptosporidiosis;  
 KW competitive inhibition; attachment; invasion; ligand binding; sporozoite;  
 KW merozoite; diarrhoea; protozoacide.  
 XX  
 XX Cryptosporidium parvum.  
 OS  
 XX  
 XX US6071518-A.  
 XX  
 XX 06-JUN-2000.  
 PD  
 XX  
 XX 12-SEP-1997; 97US-0928361.  
 XX  
 XX 13-SEP-1996; 96US-0026062.  
 XX 01-JUN-1993; 93US-0071880.  
 XX 29-MAY-1992; 92US-0891301.  
 XX 03-APR-1995; 95US-0415751.  
 XX 14-AUG-1996; 96US-0700651.  
 XX  
 XX (REGC ) UNIV CALIFORNIA.  
 XX  
 XX  
 XX Petersen C;  
 XX  
 XX WPI: 2000-422065/36.  
 XX N-PSDB; AAA61848, AAA61849.  
 XX  
 XX New GP900 protein fragments and fusion proteins of Cryptosporidium  
 PT parvum, useful for detecting the presence of the parasite, and  
 PT diagnosing or treating Cryptosporidium infections by competitive  
 PT inhibition of the function of GP900  
 XX  
 XX Claim 2; Column 61-70; 59pp; English.  
 PS  
 XX  
 XX The invention relates to the GP900 glycoprotein of the protozoan  
 CC Cryptosporidium parvum, DNA encoding it, GP900 fragments, and fusion  
 CC proteins comprising GP900 fragments. The invention also relates to the  
 CC administration of GP900 or fragments thereof to a host to elicit anti-  
 CC GP900 antibody production, and to a method of cryptosporidiosis treatment  
 CC or prophylaxis comprising administration of anti-GP900 antibodies to an  
 CC individual. Cryptosporidium parvum GP900 and GP900 fragments are able to  
 CC competitively inhibit sporozoite or merozoite attachment or invasion, and  
 CC are also useful for the generation of anti-GP900 antibodies. The  
 CC antibodies also inhibit sporozoite or merozoite attachment/invasion and  
 CC additionally inhibit the binding of GP900 ligands to GP900. GP900  
 CC proteins, fragments and antibodies may therefore be used to treat or  
 CC prevent cryptosporidiosis. Infection with Cryptosporidium is a common  
 CC cause of diarrhoea in humans and causes life-threatening diarrhoea in  
 CC immunocompromised persons. Cryptosporidiosis can be contracted from  
 CC contaminated municipal water supplies (e.g., public swimming pools). It  
 CC is also a cause of disease in animals, resulting in financial losses in  
 CC agriculture. GP900 fragments, fusion proteins and antibodies may also be  
 CC used for the diagnosis of Cryptosporidium parvum infections, and for the  
 CC detection of the parasite in the environment. The present sequence  
 CC represents a portion of the GP900 protein of the NINC isolate of  
 CC Cryptosporidium parvum.  
 XX  
 XX Sequence 1721 AA;  
 SQ  
 Query Match 7.5% Score 488; DB 21; Length 1721;  
 Best Local Similarity 22.5% Pred. No. 6,2e-17;  
 Matches 303; Conservative 128; Mismatches 469; Indels 448; Gaps 58;  
 QY 214 MGFSSNGVADICSNHIGTKG-LNDNNYPVSSSEFSYTKTC-----TSNGIQI 263  
 Db 19 infstgttdtsnmwpvstsgelkd---pnkqatigsrscgwkqysidsstgfrv 75

QY 264 -----KYONVPAGYRPFIDAYIS-----ATDVNQYTLAYTNDYTCAGSRLOSKPFTL 310  
 Db 76 dsitglptdysncp--fnpvtgnlvarstgktpntpygyvrsnel-----ktepsan 128  
 QY 311 RWTG-YKNSD-----AGSNGIVI----- 327  
 Db 129 tyagvyrsnckttcpsanctnflldpkinapnonsensfeagqifdmgskvlypyckevg 188  
 QY 328 -----VATRTVDSSTAVTTTLFPNPSVDKTKIEILOPIPTTTTTSYGVGVTS 377  
 Db 189 vkhttt 248  
 QY 378 YLTKTAPIGETATVIDVYHTTVTSEWTFITTTTTRTNPDSIDTVVQVPLNPT 437  
 Db 249 ttt 308  
 QY 438 VSTTEYWSQSFAFATTTTAPPGGDTVIIREPNNHTVTTEYWSQSFAFATTTTAPPGGT 497  
 Db 309 ttt 359  
 QY 498 DSVIIRBPNNPTVTTEYWSQSFAFATTTTAPPGGSDSVIIRBPNNPTVTTEYWSQSYA 557  
 Db 360 -----tt 401  
 QY 558 TTTTVPATPGGSDSVIIR----- 580  
 Db 402 ttttttse---tesvikpdewlekegeceakgatyvgvlgkgdriengmaftmipndd 458  
 QY 581 TTTTTEYWSQSFAFATTTT-----TAPP----- 602  
 Db 459 thvfrfkvdgntslsvrcrkagkiefdrslldftlppvaghnsclivgvsdgdklh 518  
 QY 603 ---GGNDTVIIRREP--PNH-----TVTTTEY-----WSQSFAF-----TTTVT 635  
 Db 519 vspygskdvlisapilqpselnefyvcdtctakgaihsgyqtsadfvttttakpttttt 578  
 QY 636 GPESGDTVIIREPNNPTVTTEYWSQSFAFATTTTAPPGGDTVIIREPNNHTVTTEY 695  
 Db 579 gapqgptttttgspkpttttt-----tkattttttlnp-----ilittt- 617  
 QY 696 WSQSFAFATTTTAPPGGDTVIIREPNNHTVTTEYWSQSFAFATTTTAPPGGDTVII 755  
 Db 618 --tqkpttttttkvpq-----kplattttt---lkpivtttttkattttttv--- 661  
 QY 756 EPPNPVTVTTEYWSQSFAFATTTTAPPGG-----TDTVIIYESMSSSKIST 801  
 Db 662 ----ptttttt---krdemtttttldpdlgdlleltplekmlkymlydngslld 715  
 QY 802 SSN-----DITSIIIP-----SFSRPH-YVNSTTSOLS 827  
 Db 716 ndeplpssagqgiadtlnlfpvqthkstgldpdmvglpfdpksgnlvhyptnqmsgl 775  
 QY 828 TFESSNMNTPTSISSD-----GMLSLTLVTESETTTTELICSDGKECRLSS 876  
 Db 776 vsylaakn--ltvtdetyglpidltlgypldpslfpnptgetgelfdpisdelmngti 832  
 QY 877 SGIVTNPDSNES--SIVTSTVTPTASTM-----SDLSSTDGI-----SATSSDNV 919  
 Db 833 agivsglsasesllsqksalldpdmnmvgefglpnlpatgvmipglfpseqtqfpei 892  
 QY 920 SKSGSVVTTESTVTTIQTTPNPLSSSVTSLTQLSSIPSVSES-----ESKVTFTSN--GDN 973  
 Db 893 edggi-----ippevaanaadkfklisppsvpesipekdqkdiselmydi 939  
 QY 974 QSGHDSQ-----STSTEIEIVT-----TSSTKVLPPVSSNNTDLT 1009  
 Db 940 esgrlqgvskrpipgslagdlpimktptqtdsvtqkpldptgtgipfnpp-----tghl 994  
 QY 1010 SEPTNREQPTLIS-----TTSNITEDITTSOPTGD--NGDNTSSTNPVPTVATS---T 1059  
 Db 995 lnptnnntmsssfagaykavysngikcdnvvgilpvygeitglpkpogsdipfnstgelvd 1054

QY 1060 LASASEEDNKGSHESASTSLKPSMGENSEGL---TTSTETETATTTSPTTEAPSPAVSSGTD 1116  
Db 1055 pstgkpinntagivsgkpglppledengnlfdpsetnlpidgnnqlvnpetnstvsgsts 1114  
QY 1117 VTTEP-----TDTREOPTTLSTTSKTNSELVATTQATNENGKSPSTDLTS 1162  
Db 1115 gttkpggipvngggvvpdeekdq-----adkkgdgliivptnsin----kdpvtntqy 1165  
QY 1163 SLTTGTSTASTANSELVTSQVTCGAVASANDQSHSTSVTNSNSIVSNTDPTLSQOVT 1222  
Db 1166 snlttg-----nllnpe---tgkvi pgsipgslnypsf-----ntpqct--deit 1204  
QY 1223 SSSPSTWTFIASTYDCS-GSIQHSTWL 1249  
Db 1205 gkpvdvtvgi--pydpstgeildpackl 1230

arch completed: October 3, 2002, 15:32:14  
b time: 366 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:26:33 ; Search time 16.32 Seconds  
(without alignments)  
1885.800 Million cell updates/sec

Title: US-09-715-876-8  
Perfect score: 6495  
Sequence: 1 MLQQTLLFLYLISASAKTI.....SIHQSTNLYGLITLSLFI 1260

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.AA.\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	940	14.5	1537	1 US-08-325-267A-2	Sequence 2, Appl
2	516	7.9	894	3 US-08-362-525-22	Sequence 22, Appl
3	516	7.9	894	3 US-08-971-692-15	Sequence 15, Appl
4	488	7.5	1721	3 US-08-928-361B-6	Sequence 6, Appl
5	487	7.5	1837	3 US-08-700-651-5	Sequence 5, Appl
6	487	7.5	1837	3 US-08-928-361B-5	Sequence 5, Appl
7	468.5	7.2	862	1 US-08-325-267A-4	Sequence 4, Appl
8	404.5	6.2	2035	1 US-08-046-585-5	Sequence 5, Appl
9	404.5	6.2	2035	1 US-08-393-703-5	Sequence 5, Appl
10	404.5	6.2	2035	5 PCT-US93-11721-5	Sequence 2, Appl
11	350	5.4	650	3 US-08-362-525-2	Sequence 2, Appl
12	323	5.0	907	3 US-08-783-774-2	Sequence 2, Appl
13	323	5.0	907	5 PCT-US95-04611A-19	Sequence 19, Appl
14	322	5.0	1481	2 US-08-616-844-40	Sequence 40, Appl
15	322	5.0	1481	2 US-08-599-654-40	Sequence 40, Appl
16	322	5.0	1481	3 US-08-944-868A-40	Sequence 40, Appl
17	322	5.0	1481	3 US-08-944-423A-40	Sequence 40, Appl
18	322	5.0	1481	3 US-08-944-496-40	Sequence 40, Appl
19	310.5	4.8	249	3 US-08-700-651-15	Sequence 15, Appl
20	310.5	4.8	249	3 US-08-928-361B-20	Sequence 20, Appl
21	304.5	4.7	2409	6 5180808-2	Patent No. 5180808
22	304	4.7	1203	3 US-08-351-200-2	Sequence 2, Appl
23	303	4.7	216	3 US-08-928-361B-27	Sequence 27, Appl
24	301	4.6	216	3 US-08-928-361B-8	Sequence 8, Appl
25	297	4.6	750	4 US-09-163-239A-4	Sequence 4, Appl
26	296	4.6	1848	4 US-08-296-791-6	Sequence 6, Appl
27	296	4.6	1848	5 PCT-US95-10661A-6	Sequence 6, Appl

Sequence 5, Appl  
Sequence 2, Appl  
Sequence 2, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 33, Appl  
Sequence 33, Appl  
Sequence 47, Appl  
Sequence 1, Appl  
Sequence 1, Appl  
Sequence 2, Appl  
Sequence 4, Appl  
Sequence 36, Appl  
Sequence 14, Appl

US-08-325-267A-2  
Sequence 2, Application US/08325267A  
Patent No 5585271  
GENERAL INFORMATION:  
APPLICANT: WATARI, JUNJI  
APPLICANT: TAKATA, YOSHIHIRO  
APPLICANT: OGAWA, MASAHIRO  
APPLICANT: PENITILA, MERJA  
APPLICANT: ONNELA, MAIJA-LEENA  
APPLICANT: KERANEN, SIRKKA  
TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST  
TITLE OF INVENTION: CONTAINING THEM  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
STREET: 1735 S. JEFFERSON DAVIS HWY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202

ALIGNMENTS

US-08-325-267A-2  
Sequence 2, Application US/08325267A  
Patent No 5585271  
GENERAL INFORMATION:  
APPLICANT: WATARI, JUNJI  
APPLICANT: TAKATA, YOSHIHIRO  
APPLICANT: OGAWA, MASAHIRO  
APPLICANT: PENITILA, MERJA  
APPLICANT: ONNELA, MAIJA-LEENA  
APPLICANT: KERANEN, SIRKKA  
TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST  
TITLE OF INVENTION: CONTAINING THEM  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
STREET: 1735 S. JEFFERSON DAVIS HWY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202

COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/325.267A  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP PCT/JP94/00290  
FILING DATE: 24-FEB-1994  
PRIOR APPLICATION NUMBER: JP 38871/1993  
FILING DATE: 26-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2589-023-0XPCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1537 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein





APPLICATION NUMBER: EP 92202080.5  
 FILING DATE: 08-JUL-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: EP 92203899.7  
 FILING DATE: 14-DEC-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/EP93/01763  
 FILING DATE: 07-JUL-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KOKULIS, PAUL N.  
 REGISTRATION NUMBER: 16,773  
 REFERENCE/DOCKET NUMBER: 213289/77020(V)  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 861-3000  
 TELEFAX: (202) 822-0944  
 TELEX: 6714627 CUSH  
 INFORMATION FOR SEQ ID NO: 22:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 894 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-362-323-22

Query Match 7.9%; Score 516; DB 3; Length 894;  
 Best Local Similarity 26.3%; Pred. No. 4.5e-22;  
 Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

QY 437 TVSTTEYWSQSFATTTTAPPGGTDVIREPPNHTVT--  
 DB 53 TYSNAAWYAGYASKTKLSGGGTDISIDNIPCVSSSGTFPCQEDSYGNWCKMGMA 112  
 QY 478 -----EYWSQS-FATTTTAPPGGTDVIREPPNHTVT--  
 DB 113 CSNSOGIAYWSTDLFGFYTT-----PTNVTLEMTGYFLPPQTGSGYTFKEA 157  
 QY 522 TT-TTATAPPGGTDVIREPPNHTVTTEYWSQYATTTTAPPGGTDVIREPPNHT 580  
 DB 158 TVDSSAIVSGGATAFNCCAOQPPITSNF-----TIDGKPMGGS-----LPPNI 204  
 QY 581 TVTTEYWSQY-----ATTVTAPPGGT-----DTVI 609  
 DB 205 EGTVMYAGYIPMKVYVSNVAVSNGTLPISVTLPGDTPGVYVYFDDLSQSNCT 264  
 QY 610 IREPPNHT-----TVTTEYWSQSFATTTTAPPGGTDVIREPPNHTVT--TT 656  
 b 265 VPDPNSAVSTTTTTEPWTGTFTSTEMTIVTGTNGVPTDETIVIRTPSEGLIST 324  
 QY 657 TEYWSQYATT-----TTTAPPGT-----DTVI 609  
 DB 325 TPTMTGTFTSTSTEMTIVTGTNGVPTDETIVIRTPSEGLISTTTEPWTGTFTSTEM 384  
 QY 704 TTVTAPPGGTDVIREPPNHTVTTEYWSQYATTTTAPPGGTDVIREPPNHT--T 761  
 DB 385 TVTGTNG-----QPTDET-----IVIRTPSEGLV 411  
 QY 762 TVTTEYWSQSFATTTTAPPGGT-----DTVI 609  
 DB 412 TTTTEPWTGTFTSTSTEMTIVTGTNGVPTDETIVIRTPSEGLISTTTEPWTGTFTSTEM 811  
 QY 812 SFR-----PHVNSTSDLSSTFESSNMPTTSSDGMLLSSTLVLTETSTTELCSDG 867  
 DB 472 S-SRITTPFNGSVTSSSVSSVTSLSFTSPVI--SSSVISSTTSTSTSES 528  
 QY 868 KCSRSLSSSGVTPNDSNNESSIVTSTVPTASTMSDSLSTSDGISTSDNYSKSGSVT 927  
 DB 529 SKSVIPSTSSSTSGSESTSS-----AGSVSSSFSISSESKSPSTYSSSLP 576  
 QY 928 TETSVTITQTPNPPLSSVTSITOLSSIPSVSESKYVTFSTNCDNQSHDSQSTTEI 987  
 DB 577 LVTSATTSQETASLPPATTTKT-----SEQTTLTVT-----TSCS 613

QY 988 EIVTSTKVLPPVYSSNTDLTSEPTNTREOPTTLSTTSNSITEDITTSOPTGDCNDVTS 1047  
 DB 614 HVCES-----ISPAIVSTATVTVSGVTVTEYTWCPISTTE-----TTKOTKGT--EQTT 662  
 QY 1048 STNPVPTATSTLASAEEED--NKGSGHESASTSLKPSMGENSEGLTT-----STELVAT 1100  
 DB 663 ETRKQTVTVI---ISSCESDVCSTASPAIVSTSTATINGVTVTEYTWCPISTTESRQOT 719  
 QY 1101 TSPT-----EAPSPA-VSSGTD-----VTTEPTDTRQOPTTLSTTSKTNSELVAT 1144  
 DB 720 TLTVTVSCSVCSETASPAIVSTATVTVTVTPWRPQTANEEVSVKMSATGET 779  
 QY 1145 TQATNENGKSPSDDLSTSLTGTSTASANSSELVTSVGTGG-----AVASASNDQSHS 1199  
 DB 780 T--TNTLAAETTTNTVAETITNCA--AETKTVVTVSSLSRSHNAETOTASATDVIGHS 834  
 QY 1200 TSV-----TNSNSIVSNTPQTLTSCQVTSSTP-----STNTFFIASTYDGS 1241  
 DB 835 SSVSVSVSETGNTKSLTSSGLSTMSQPRSTPASSWGVYSTASLEISTYAGSAT 887

RESULT 3  
 US-08-971-692-15  
 ; Sequence 15, Application US/08971692  
 ; Patent No. 6114147  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: Immobilized proteins with specific binding  
 ; NUMBER OF SEQUENCES: 40  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/971,692  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; INFORMATION FOR SEQ ID NO: 15:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 894 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-971-692-15

Query Match 7.9%; Score 516; DB 3; Length 894;  
 Best Local Similarity 26.3%; Pred. No. 4.5e-22;  
 Matches 251; Conservative 126; Mismatches 310; Indels 266; Gaps 43;

QY 437 TVSTTEYWSQSFATTTTAPPGGTDVIREPPNHTVT--  
 DB 53 TYSNAAWYAGYASKTKLSGGGTDISIDNIPCVSSSGTFPCQEDSYGNWCKMGMA 112  
 QY 478 -----EYWSQS-FATTTTAPPGGTDVIREPPNHTVT--  
 DB 113 CSNSOGIAYWSTDLFGFYTT-----PTNVTLEMTGYFLPPQTGSGYTFKEA 157  
 QY 522 TT-TTATAPPGGTDVIREPPNHTVTTEYWSQYATTTTAPPGGTDVIREPPNHT 580  
 DB 158 TVDSSAIVSGGATAFNCCAOQPPITSNF-----TIDGKPMGGS-----LPPNI 204  
 QY 581 TVTTEYWSQY-----ATTVTAPPGGT-----DTVI 609  
 DB 205 EGTVMYAGYIPMKVYVSNVAVSNGTLPISVTLPGDTPGVYVYFDDLSQSNCT 264  
 QY 610 IREPPNHT-----TVTTEYWSQSFATTTTAPPGGTDVIREPPNHTVT--TT 656  
 DB 265 VPDPNSAVSTTTTTEPWTGTFTSTEMTIVTGTNGVPTDETIVIRTPSEGLIST 324  
 QY 657 TEYWSQYATT-----TTTAPPGT-----DTVI 609



QY 802 SSN-----DITSIP-----SFSRPH-YVNSTSDLS 827  
Db 716 NDEPIGSOAGIADTSLNLPVQTHKSTGLPDMVGLPDPKSGNLVHPVNTQTHSGLS 775  
QY 828 TFESSMNTPTSSD-----GMLSSSTLVTESEITTELICSDKCSRLSS 876  
Db 776 VSYLAANK-LVVDRTYGLPDTLTGYPLDVPVSLIPNPFETGEFLPDSLEIMN-GRI 832  
QY 877 SGIVNTPDNSHES--SIVTSTVPTASTM-----SDLSSTDGI-----SATSDNV 919  
Db 833 AGIVSGISASESLLSOKSALIDPATNMVGFEGGLLPATGVMVIFGLPSEQTQFSPEI 892  
QY 920 SKSGSVTTTETSVTIOTPNPLSSSVTSLTOLSSIPSVSES-----ESKVTFTSN--GDN 973  
Db 893 EDGII-----IPEVAANADKFKLSIPSPESIPESIPESIPESIPESIPESIPES 939  
QY 974 QSTHDSQ-----STSEIEIVT-----TSSTKVLPPVSSNTDLT 1009  
Db 940 ESGRLICQSKRIPGSGTAGLNPIMKTPQDSVTCGKPIDPTGLPKNP-----TGH 994  
QY 1010 SEPTNREQPTLS-----TTSNITEDITTSOPTGD--NGDNFSTNVPVATVS-----T 1059  
Db 995 INPTNNMTHDSSFAGAYKAVSNGIKTNVYGLPVGEITGLPKDPSGDIPLNPNSTGELVD 1054  
QY 1060 LASASEDNKSGSHESASTSLKMGSENGSL---TTSTEIEATTSPTAPSPAVSSGTD 1116  
Db 1055 PSTCKPINNSTAGIVSGKPLPIEDENGLNLPDPSTNLPIDGNNQLVNPETNNTSVSGSTS 1114  
QY 1117 VTTEP-----TDTREQPTLTSTSKTNSELVATTQATNENGKSPSTDLTS 1162  
Db 1115 GTAPKPGIPVNGGVDPDEAKDQ-----ADKGDGLIVPPTNSIN-----KDPVTNTOY 1165  
QY 1163 SUTGTGASTANSALVTSVGTGGVAGASANDQSHSTSVTNSNISVNTPTTLTQQVTT 1222  
Db 1166 SNTTG-----NINPE---TGKVIPTGSLGSLNYPSP-----NTPQQT--DEIT 1204  
QY 1223 SSSPSTWTFIATYDGS--GSIQHSWL 1249  
Db 1205 GRPVDVTGL--PYDPSTGEIIDPATKL 1230

## RESULT

US-08-700-651-5  
; Sequence 5, Application US/08700651B  
; Patent No. 6015882  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN  
; APPLICANT: LEECH, JAMES  
; APPLICANT: NELSON, RICHARD, C.  
; APPLICANT: GUT, JIRI  
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
; TITLE OF INVENTION: INFECTIONS  
; FILE REFERENCE: 480.19-4(HV)  
; CURRENT APPLICATION NUMBER: US/08/700.651B  
; CURRENT FILING DATE: 1997-08-14  
; EARLIER APPLICATION NUMBER: 08/415,751  
; EARLIER FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO. 5  
; LENGTH: 1721  
; TYPE: PRT  
; ORGANISM: Cryptosporidium parvum  
US-08-700-651-5

Query Match 7.58; Score 487; Db 3; Length 1721;

Best Local Similarity 22.38; Pred. No. 5.2e-20;

Matches 293; Conservative 113; Mismatches 411; Indels 498; Gaps 52;

QY 297 TCAGSLQSKPFTLRWTGYKNSDAGSNGIVIVATRTVTDSTTAVTT-----LPFNPSVD 351

Db 52 TISGRS-----CGMK-----QGYSDSDSTGFRVDSITGLPTDPSYNCFPNPVTG 96  
QY 352 KTKTIELOPIPTT-----TITT-----SVGVVTSYLTAKTA-PIGETATVIVD-- 394  
Db 97 NLVSRSTGRTIPNTYAGVYRSNETKTPESANTYAGVYRSNETKTPESANTFELLVDPK 156  
QY 395 -----HTTTVTSEWGTITTTTTTNTDTS 423  
Db 157 INAPCENSESPGOIFDMGSKVYIDYKCVGVKHTTTTTTTTTTTTTTTTTTTTTTTT 216  
QY 424 IDTVVVOVLPNPTVSTTEYWSQSFATTTVTAPPGCTOTVIRPPNPHVTTTEYWSOS 483  
Db 217 TTT-----TT 266  
QY 484 FATTVTAPPGCTDSVIRPPNPTVTTTEYWSQSFATTTVTAPPGCTDSVIRPPN 543  
Db 267 TTTT-----TTKPT 309  
QY 544 PTVTTEYWSQSFATTTVTAPPGCTDSVIRPPNPTVTTTEYWSQSFATTTVTAPPC 603  
Db 310 TTTT-----TTT 355  
QY 604 GTDVIIRPPNPTVTTTEYWSQSFATTTVTGPPGCTDVIIRPPNPTVTTTEYWSOS 663  
Db 356 -TTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 400  
QY 664 YATTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 671  
Db 401 -ATTTT-----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 459  
QY 672 -----TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 671  
Db 460 HVREFRKVDGNTISVRCRKGAGKLEFPDRSLDFTIPPVAGHNSCSIINGVSGDGKIHV 519  
QY 672 APPGETDVLIRPPNPTVTTTEYWSQSYA-----TTTTTTTTTTTTTTTTTTAKPT 704  
Db 520 SPYSGKDVLSIAP-----IQPSELFNEVYCDCTAKYGAHSGYQTSADFVTTTAKPT 575  
QY 705 TVTAPGETDVLIRPPNPTVTTTEYWSQSYA-----TTTTTTTTTTTTTTTTTTAKPT 764  
Db 576 TTTGAPCQPTTTTSGPSKPTTTT-----TKATTTTTLNLPIITTT--OKPTTTTT 627  
QY 765 TEYWSQSFATTTVTAPPGCTDVIIRPPNPTVTTTEYWSQSFATTTVTAPPGCTDVIIRPPNPTVTT 824  
Db 628 KVPKPPPIATTTTTLKPIVTTT-----TKATTTT-----TTTVP-----TTTT 666  
QY 825 DLSTFSSSMNTP-----TSISSDGLLSSTTLVTESEITTELICSD---GKECS 871  
Db 667 TTKRDEMTTTTLPDIDGITEITPIEKMLDKVTRMIDYNSGLLSDNDEPIPGSQAG 726  
QY 872 RLS-----SSSGIVTNP-----DSNESSIV-TSTVPTASTMSDLSSTGISA 913  
Db 727 QIADTSLNLPVQTHKSTGLPDMVGLPDPKSGNLVHPVNTQTHSGLSGLVLAANKLIV 786  
QY 914 TSSDNVSKSGSVTTTETS-----VTTIOTPT-----NPLSSSVTSLTOLSSIPSVSESE 962  
Db 787 DTDIVY---GLPIDLTGYPDVPVSLIPNPFETGEFLPDSLEIMNAGTIVSGISASE 843  
QY 963 S-----KVTFSTGNDNQ-----SCTHDSQSTSTEIEIVTTSSTKVL 998  
Db 844 SLLSQKSALIDPATNMVGFEGGLLPATGVMVIFGLPSEQTQFSPEI-----DGGII 898  
QY 999 PP-VVSSNTD--LTSEPTNREQPTLTSTNSITE---DITTSQPTCD-----NGDWT 1046  
Db 899 PPEVAANADKFKLSIPSPESIPESIPESIPESIPESIPESIPESIPESIPESIPES 958  
QY 1047 SSTNPTVATST-----LASASEDNKSGSHESA-----STS 1079  
Db 959 GDLPINMKPTQDTSVTCGKPIDPTGLPKNPPTGHLINPTNNMTDSSFAGAYKAVSNG 1018  
QY 1080 LKPS-----MGSENGSL-----TTSTEIEATTSPTAPSPAVSSGT----- 1115

```
Db 1019 IKTDNYGLPVGEITGLPKDPGSDIPFNSTGTBLVDPSTGKPIINNSTAGIVSGKPLPPI 1078
Qy 1116 -----DVTE-PTDREQ---PTLSTSKTNSLVAATQATNENG----- 1153
Db 1079 EDENGLFDPSTNLDPIDGNOLVNPETNSGSGTKPKGPVNGGVVPDEAKD 1138
Qy 1154 -----KSPSTDLSTSSLTGTSASTSANSSELVTSVGGAVASND 1195
Db 1139 QADKGDGLVPPNTSINKDPVTNTQVNTTG---NIINPE---TGKVIPLGSLGSLNY 1191
Qy 1196 QSHSTVTSNSIVSTPQTLLSQOVTSSSPSTNTFIASLYDGS-GSIIOHSTWL 1249
Db 1192 PSF-----NTPQPT--DEITGKPVDTVTGL--PYDPSGTGIIDPATKL 1230

RESULT
US-08-928-361B-5
; Sequence 5, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1837 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-5
```

```
Query Match 7.58; Score 487; DB 3; Length 1837;
Best Local Similarity 22.88; Pred. No. 5.7e-20;
Matches 316; Conservative 146; Mismatches 487; Indels 440; Gaps 61;

Qy 136 VGTGSGSTDLSDSKFTAGNTVTFN-DGDKDISIDVFEKST---VDP-SAYLY--ASR 188
Db 69 IDSTGKKFSPYTGKHADASTTSAYSAPFELDVS-GVPIENTRMVDPVSLUMFDNSTG 127
Qy 189 VM-PSLUNKVTLFVAPQCENGYTSCTGMFSSNGDVAIDCSNIHIGITKGLDWNVPVS 246
Db 128 VMYDPNTNISLEGSIAIRSESCIVSELNFTTGTGTTDTSMMNPVSITSG----- 178
```

```
Qy 247 SESFSYKTKCTSGNGIQKQNVNYPAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLOSK 306
Db 179 -----ELKDPNKQA-----TISGRS--- 194
Qy 307 PFLRWGTYKNSDAGSNGIVIVATRTVTDSTAVTT-----LPFNFSVDKTKTIEILOP 361
Db 195 -----CGWK-----QGYSDSSTGFRVDSITGLPTDPYPCNCFNP-----VTGNLVR 237
Qy 362 IPTTTTTSYGVGVTTSYLTKA-PIGETATVIVD----- 394
Db 238 STGKTIPNTVAGYVRSNETKTEPSANTNELLVDKPINAPCNSENSEFEQVIFDMGSKVY 297
Qy 395 VPY-----HTTTVTSEWGTITTTTTRNPTDSIDTVVQVPLNPNTVSTTEYWSOS 447
Db 298 IPYTKCVGVKHTTTTTTTT-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 356
Qy 448 FATTTVTAPPDGTDTVIIRPPNHTVTTEYWSQSFAATTTVTAPPDGTDSVIREPPN 507
Db 357 TTTTTTTTTTTTTTTTTTKKPTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 401
Qy 508 PTVTTTEYWSQSFAATTTVTAPPDGTDSVIREPPNPTVTTEYWSQSAYATTTTVPAPG 567
Db 402 TTTTTTTTTTTTTTTTTTKKPTTTTTT-----TTTTTTTTTTTTTTTTTTTTTKPT 457
Qy 568 GTDSVIREPPNHTVTTEYWSQSAYATTTVTAPPDGTDTVIIRPPNHTVTTEYWSOS 627
Db 458 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-----TTTKKPTTTTTTTTTTKK 511
Qy 628 FATTTVTGPPSGDGTVIIRP----- 649
Db 512 PTTTTTATTTTTSETESVIKDEMCWLEKNGEAKGATVGVIGKDGRIENGMAFTMI 571
Qy 650 PN-PTVTTTEYWSQSAYATTTI-----TAPP----- 674
Db 572 PNDQTHVRFREKVKDVGNTISVRCGKAGKLEFPDRSLDFIPPVAGHNSCSIIIVCSGG 631
Qy 675 -----GETDVLIREPPNHTVTTEYWSQSAYT-----TTVT 707
Db 632 GKIHSVPGSKDVSLSAP-----IQCELENEYCDCTAKYGAHSGYOTSADFVTTT 687
Qy 708 APPGETDVLIREPPNHTVTTEYWSQSAYATTTVTAPPDGTDTVIIRPPNPTVTTEY 767
Db 688 AKPTTTTGTAGQP-----TTTTGSPSKPTTTTT--KATTTTIL---NPIITTT-- 734
Qy 768 WSQSFAATTTVTA---PBGCTDVTIYESMSSKISTSNDSITSIIPFSRPHYVNSTS 824
Db 735 -TQKPTTTTITKVPCKPPIATTTTTLKPIVTITTKATTTTTT-----VPTTT 783
Qy 825 DLSTFESSMNTP-----TSISDGMLLSSTLTATESETTELICSD---CKECS 871
Db 784 TTKDEMTTTPPLPDIGDIEITPIPIEKMLDXYRMIYDYNGLLDSNDEPIQSQAG 843
Qy 872 RLS-----SSSGIVTNP-----DSNESSIV-TSIVPTASTMSDLSSTDGISA 913
Db 844 QIADTSLNLPVQTHKSTGLPIDPMVGLPDPKSGNLVHPYTNQIWSGLSVSLAKNLIV 903
Qy 914 TSSDNVSKSVVTTETS-----VTTIOTTP-----NPLSSSVTSLTQLSSIPSVSESE 962
Db 904 DTDETY---GLPIDLTGYPLDPVSLPFPNPTGELFDPISDEIMNGTIAAGIVSGISASE 960
Qy 963 SKVFTFTSN-----GD-----NOSGTHDSQSTSTEIEIVTTSKVL 998
Db 961 SLLSQKAPDIPATNMVGEFGLLNIPATCMIPGSLGSEQTFPSPEIE-----DGGII 1015
Qy 999 PP-VVSNQD--LTSEPTNTREQPTTLSTTSNITE---DITTSQPTGD-----NGONT 1046
Db 1016 PPEVAANADAKFKLSIPPSPVSPESPEKDKIDSISLMDYIESGLRGIGQVSKRPIPGSIA 1075
Qy 1047 SSTNPVTATSTLASAEEEDNKGSHESASTSL--KPSMGENSEGLTTSIEATTTPT 1104
Db 1076 GDLNPIMKPTQT-----DSVTGKPIDPTTGLPFPNPTGH---LINPTNNMTDSSFA 1125
```

QY 1105 EAPSAVSSG--TD-----VTEPTDTRPTTLSTRTSKTNSLY--ATTOATN----- 1149  
Db 1126 GAYKAVSNGIKTNYVGLPDEITGLPDPVSDIPFNSITGELVDPSTGAPINNYTAGI 1185  
QY 1150 -----ENGCK-SPSTD-----TSSLTTC-TSASTSANSSELYTSGS 1183  
Db 1186 VSGKRGLPPTEDENGLPDPSTKLPIDGNQLVNPETNSVSGTSGTKPKGPIVNG- 1244  
QY 1184 VTGGAVA-SASNDQSHSTS-----VNSNSIVNTPTQTLSCQVTSSTPTNFTASTYDG 1238  
Db 1245 --GGVYDEKADQADKDGGLIVPPTNSI-----NKDPVTNTQYSNT--- 1285  
QY 1239 SGSIHQHST 1247  
Db 1286 TGNINPST 1294

RESULT 7  
S-08-325-267A-4  
Sequence 4, Application US/08325267A  
Patent No. 5585271  
GENERAL INFORMATION:  
APPLICANT: WATARI, JUNJI  
APPLICANT: TAKATA, YOSHIHIRO  
APPLICANT: OGAWA, MASAHIRO  
APPLICANT: PENTTILA, MERJA  
APPLICANT: ONNELA, MAIJA-LEENA  
APPLICANT: KERANEN, SIRKKA  
TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST  
NUMBER OF SEQUENCES: 7  
CONTAINING THEM  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUBAUER  
STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/325,267A  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP PCT/JP94/00290  
FILING DATE: 24-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 38871/1993  
FILING DATE: 26-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2589-023-0XPCT  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 862 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-325-267A-4

Query Match 7.2%; Score 468.5; DB 1; Length 862;  
Best Local Similarity 26.3%; Pred. No. 2.3e-19;  
Matches 231; Conservative 120; Mismatches 297; Indels 229; Gaps 40;

QY 545 TVTTEYWSQSVAATTTTVPAPGQSDSVIIREPPNHHVTTT----- 585  
Db 53 TYSNAATYAGYASKTKIGSVGGQTDISIDYINPCVSSSGTFFCPOEDSYGNMCKGMGA 112  
QY 586 -----EYWSQS-----YATTTT-----APPQGT-----DTVII----- 610  
Db 113 CSNSQGIAYMSDLEFGFTYTPNTLEMGVFLPQGTGYTFKFAVDDSDAILSVSAGATA 172  
QY 611 -----REPP-----NHTVTTEYWSQSFAATTTTVPGPSCTDVIIRE-----PNPTVT 656  
Db 173 FNCAOQOPITSTNFTIDGKPMGSL-----PPNIEGTVMYAGYYPKVVYSN 224  
QY 657 TEYWSQSVAATTTTITAPGET-----DTVLIREEPPNH-----TVTTTE 694  
Db 225 AVSMG--TLPIISVTLPGDITVSDDEFGYVVSFDDLSQSNCTVDPDSNYAVSTTTTTE 281  
QY 695 YWSQSVAATTTTITAP--PCETDTVLIREEPPNH--VTTEYWSQSVAATTTT 741  
Db 282 PMTGTFTSTSTMTTGTNGVPTDVTIVIRPTTASTIITTEPMTGTFTSTSTMTT 341  
QY 742 VTAPPG-GTD--TVIIREPPNP--TVTTEYWSQSFAATTTTVPAPGQGT-----DTVII 790  
Db 342 VTGTNGQPTDVTIVIRPTSEGLVTTTPEPMTGTFTSTSTMTTGTNGVPTDVTIV 401  
QY 791 YE---SMSSKISTSSNDIITSIPFSR---PHYVNSTTDLSTFESSMNTPTSISS 842  
Db 402 VKTPTTAISSLSLSSSGQITSSITS-SRPIITFPYPSNGTSVISSSVISSTSLSTS 460  
QY 843 DGMILLSTTLVTESETTTTELICSDGKCSRSLSSSGSIVTPNPSNESSIVTSTVPTASTMS 902  
Db 461 SPVI---SSSVISSTTTSTIFSESSKSVIPTSSSTSGSESEWS-----A 506  
QY 903 DLSLSTDGISATSSDNVSKGVVTTTSTVTTIOTNPPLSSSVTSLTQLSIPSVSESE 962  
Db 507 GSVSSSFISSESSKSPYSSSLPLVTSATTSOETASSLPATTTKT-----SEQT 558  
QY 963 SKVTFTSNGDNQSGTHDSQSTSEIIVTTSKVLPPVYSSNTDLTSEPTNTREOPTTL 1022  
Db 559 TLVTV-----TSCSHVCTES---ISPAIVSTATVTVSGVTEYTTWCPI 600  
QY 1023 STTNSITEDITTSQPTGDNGDNSTSNPVPVATSTLASAED--NKSGSHESASTSL 1080  
Db 601 STTE-----TTKQTKCTT-EQTETTKQTTVT---ISSCESDVCSTASPAIVSTST 649  
QY 1081 KPSMGENSEGLTT-----STEIETATTSPT-----EAPSPA-VSSGTD-----VTT 1119  
Db 650 ATINGVTEYTWCPISITTESRQOTLVTVTSCESGVCSSETASPAIVSTATVNDVTV 709  
QY 1120 EPTDTRPTTLSTRTSKTNSLYVATQATNENGKSPSTDLTSSLTGTSTASTSANSSELY 1179  
Db 710 YTHRPOTANEESVSKNSATGETT--TNTLAAETTTNTVAEITVNGA---AETKV 764  
QY 1180 TSGSVTCG-----AVASANDQSHSTS-----TNSNSIVNTPTQTLSCQVTSSTSP--- 1226  
Db 765 VTSSLRSNNAETOTASATDVIGHSSSVSVSEVETGNTKSLTSGUSTWSQOPRSTPASSM 824  
QY 1227 ---STNTFTASTYDGSIIQHSITWLYGLITLLSLFI 1260  
Db 825 VGYSTASLEISTYAGSANSLLAGSLGVFIASLLAI 861

RESULT 8  
US-08-046-585-5  
Sequence 5, Application US/08046585  
Patent No. 5453362  
GENERAL INFORMATION:  
APPLICANT: Lamarco, Kelly  
APPLICANT: Wilson, Angus  
APPLICANT: Herr, Winship  
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:  
TITLE OF INVENTION: HOST CELL FACTOR  
NUMBER OF SEQUENCES: 15

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT  
;; STREET: 4 Embarcadero Center, Suite 3400  
;; CITY: San Francisco  
;; STATE: CA  
;; COUNTRY: USA  
;; ZIP: 94111-4187  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/046,585  
;; FILING DATE: 12-APR-1993  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Osman, Richard A  
;; REFERENCE/DOCKET NUMBER: 36,627  
;; TELEPHONE: (415) 781-1989  
;; TELEFAX: (415) 398-3249  
;; TELEX: 910 277299  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2035 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: peptide  
;; US-08-046-585-5

Query Match 6.2%; Score 404.5; DB 1; Length 2035;  
Best Local Similarity 21.9%; Pred. No. 3,7e-15;  
Matches 319; Conservative 165; Mismatches 506; Indels 467; Gaps 69;  
QY 14 IASAKTITGVDFSNLSWNAANAFKPGVPTWNAV-LGWSLDGTSA-NPGDFTLNM 71  
DB 537 ICGSPQMSGH---AALAAAAATKIPPSAPTIVLSVPAGTIVTKMAVTPG---TTTL 589  
QY 72 PCVKFYTSQTSVDLTADGVKATCQFYSGEFTFTLTCTVNDALKSSIKAFGTPLP 131  
DB 590 PATVKVASSPMV-----SNPATRMLKTAQAQVGTSSV---ATNTSTRP 631  
QY 132 IAFNVGVTGSTDLEDKCFEAGTNTVTFTNDGDKDISDVEFEKSTVDPSAVLYASRVP 191  
DB 632 I-ITVHSGTNTVAQAQVTTVVGGVT-----KTTIL-----VKSPISVPGGSALIS 678  
192 SLNKVTTLFVAPQENGVTSGTMGFSNCGDAIDCSNIHIGITKGLNDWNYPVSSSEFS 251  
DB 679 NLGKVMNV---QTRPVQTSATVC-QASTGPVT-----QIIQTKG-----714  
QY 252 YTKCTCSNGIOIKYONPAGYRPFIDAYISATDVNOYTLATNDYTCAGSRLOSQPFPLR 311  
DB 715 -----PLAG---TILKLVTSADGKPTVIIITTOQACAG---TKPTIL- 751  
QY 312 WTGYKNSDAGSGIVVATTTVDSTTAVTLPLPNPSVDK-----TKTIELQIP-- 363  
DB 752 -----GISVSFSTKGTGIIKIPMSALITQAGATGVTSSPGKSPITII 799  
QY 364 TTTTTSYGVVTSYLT-----KTAPIGETATVIVDVPYHTTTVT 404  
DB 800 TTKVMTSGTGAFAKIIATVPKIATGHGQGVQVVLKAP-GQPTILRTVPMGGVRLVT 858  
QY 405 SEWGTITTTTTRTPTSDITVVVQVPLNPNTVSTTEYWSQSFAITTTVTPPGGDTV 464  
DB 859 -----PVTVSANRP--AVTLVVKGTGVTTLGTVT-----GIVTSLAGAGHST- 902  
QY 465 IIREPPNHVTTTEYWSQSFAITTTVTPPGGDTSVIIREPPNPTVTTTEYWSQSFAIT 524  
DB 903 -----SASLATPITL-----GTIATLSSQVINPTAITS-----AAQT 936

QY 525 TVTAPPGGTSVIREP-PNPTVTTTEYWSQSATTTTTPAPPGTSDVIREPPNHTV- 582  
DB 937 TLTAAGGLTPTITMQPVSQPT-----OVLITAPSG-----VEAOPVHDL 978  
QY 583 -----TTTEYWSQSATTTTTPAPG---GTDVIREPP--NHTVTTTEYWSQSFA 630  
DB 979 VSILASPTTE---OPTATVTIADSGQDVGQVTVLVCNSPPCETHETGT---NTAT 1030  
QY 631 TTTVT--GPPSG-----TDTV-----TTTTEYWSQSFAITTTTTPAPG--- 1090  
DB 1031 TTVVANLGGHPQPTQVFCVCDRQEAASLVSTVVGQNGSVYRVCSNPPCETHETGT 1090  
QY 665 ATTTT-----ITAPPGET-----DTVIREPPNH-----T 689  
DB 1091 ATTATSNMAGHGCNSPPCETHETGTNTTATTATMSSVGNHQRDARRACAAAGT 1150  
QY 690 VTTTEYWS-----QSYATTTVT-----AP-----PGETDTVLIR 719  
DB 1151 VATGALEAAOGSKSQCTROTTSATSTMTVMATGAPCSAGPLLGPMSAREPGGRS 1210  
QY 720 EPP-----NHTVTTTEYWSQS-----ATTT 740  
DB 1211 LAPSSKVRLLSSPSIKDLPAGRHSHAVSTAAMTRSSVAGEPMAVPCESLOGGSP 1270  
QY 741 TVTA-----PPGGTDTVIREPPNPTVTTTEYWSQSFAITTTTTPAPPGTDT 1270  
DB 1271 TVTALEALLCPSATVTQVCSNPPCET-----HETGTTNTATTNAGSAQRCV 1323  
QY 796 SSKISTSSNDITSI-----PSFSRPHYVNSTTSDLSSTFE-SSSMNTPTSIS 842  
DB 1324 THETGTTTATTATNSNGTGQPEGGOQPPAGRCPTHQTTTGTGTTMSVSGALLP 1383  
QY 843 -----DGMLLSSTTLVTESETT-----TELICSDGKCSRLLSSSSGIVTNP-- 888  
DB 1384 HRTVESGLEVAAPSVTPQAGTALLAPPTQVCS-----NPPCETH- 1426  
QY 889 SVTSTVPTASTSDLSLSDGTSATSSD-----NVKSGSVVTTSTSTVTTIOT 943  
DB 1427 ---TGTHATTATVTSNMNSQDPPPPADOGESVTOGDSVNITSSAITT---TVS 1481  
QY 944 SSVTSLTQLSISVSSESKVTFTSNGDMQS--GTHDSQSTSTEI-----EIVT 996  
DB 1482 RAVTTVTQSTVPVCPSPVPPPELQVSPGQQLPPROLLOSASTALMGESAEVLSA 1541  
QY 997 VLPVPVSNLDTLSEPTNTREOPTTLSTNSITDITTSOPTGNGDNTSSTNPVPTVA 1056  
DB 1542 ELPAADVLS--TGEPSGQE-----SAGSAVATVVVQ-----PPPTQS 1580  
QY 1057 TSTLASAEE---DNKSGSHESASTSLKPSMGNSGLTSTTEIATTTSTPEAPS 1113  
DB 1581 EVDQLSLPQELMAEAQAGTTILMTGLTP---EELAVTAAEAQAQAATEEAQA 1637  
QY 1114 GTDV-----TTEPTDTRQPTTLSTTSKTNSELVATQATNENGKSPSTDLT 1166  
DB 1638 VLOAAQAVMGTGEPMDTSEAAVT---TQAEI-----CHLSAEGOGQATPI 1688  
QY 1167 GTSASTSNSELVTSQS-----VTGGAVASANDQSHSTSVTNSIVSNTPOTT 1216  
DB 1689 QELAAVQOOQLOQAQAOQHHLHLPTEALAPADSLNDPAIESNCL--NELAG 1746  
QY 1217 LSQVTSSTSTWTFIA 1233  
DB 1747 LPSTATSLAPSMTEVA 1763

RESULT 9  
US-08-393-703-5  
; Sequence 5, Application US/08393703  
; Patent No. 5585239  
; GENERAL INFORMATION:  
; APPLICANT: Lamarco, Kelly  
; APPLICANT: Wilson, Angus

APPLICANT: Herr, Winship  
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:  
NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HONBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/393,703  
FILING DATE: 24-FEB-1995

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-57503-2/RAO  
TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249

TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-393-703-5

Query Match 6.28; Score 404.5; DB 1; Length 2035;

Best Local Similarity 21.98; Pred No. 3.7e-15;  
Matches 319; Conservative 165; Mismatches 506; Indels 467; Gaps 69;

Qy 14 IASAKITGVDFNSLWSAANYAFKGGPYTNNAV-LGWSLDGTS-A-NPGDTFTLAW 71  
Db 537 IGSPPQSMGM-----AALAAARATQKIPSSAPTVLSVPAGTIVTKMAVTPG---FTTL 589  
Qy 72 PCVFKYTSQTSVDLTADGVKYATCQFYSGEEFTFTSLCTVNDALKSKAKFGVTPL 131  
Db 590 PATVKVASSPMV-----SNPATRMLKTAQAQVGTSSV---ATNTSTRP 631  
Qy 132 IAPNVGGTSGSDLEDSKCFAGTNVTVNDGDKDISDVEFEKSTVDPSAYLYASRVMP 191  
Db 632 I-ITVHKSGTVAQAQVTVTVVGGV-----KTITL-----VKSPISVPGGSALIS 678  
Qy 192 SLUNKVTLFVAPQENGYSGTGTMGFSNGDVAIDCSNIHIGITKGLNDWNPVSSSEFS 251  
Db 679 NLGKVMNV---QTKPVQTSVAVTQ-QASTGPVT-----QIIQTKG----- 714  
Qy 252 YTKTCTSGIQIKYQNVAGYRPFIDAYISATDVNQYTLAYTNDYTCAGSRLOQKPFLLR 311  
Db 715 -----PLPAG---TILKLVTSADGKPTTIIITTTQASGAG---TKPTIL- 751  
Qy 312 WTGYKNSDAGSNGIVAVTATVTDSTTAVTTLFPNPSVDK-----TKTTEILOQIP- 363  
Db 752 -----GISSVSPSTTKPGTTTIKTIPMSAIIITQAGATGVTSPPGKSPITII 799  
Qy 364 TTTTITSVGVTSYLT-----KTAPIGETATVIVDPYHTTTT 404  
Db 800 TTKVMTSGCAPAKIITVPRKIATGCHGQGVQVVKCAP-GQPGTILTPMGVRLVT 858  
Qy 405 SEWTCGTTTTTTRNTDSDIDVNVVQVPLNPNTVSTTEYWSQSFATTTVAPGGGTDV 464  
Db 859 -----PVTVSAVKP--AVTTLVVKTKGCTVTLGTVT-----GTVSTLAGAGGHS- 902

Qy 465 IIREPPNHTVTTEYWSQSFATTTTVPAGGTDSDVIIREPPNHTVTTEYWSQSFATTT 524  
Db 903 -----SASLATPITL-----GTIATLSQVINPALTIVS-----AAQT 936  
Qy 525 TVTAPPDGTSDVIIREP-PNFTVTTEYWSQSYATTTTVPAGGTDSDVIIREPPNHTV- 582  
Db 937 TLTAAGLGTPTITMQPVSQPT-----QVTLITAPSG-----VEAQPVHDL 978  
Qy 583 -----TTTEYWSQSYATTTTVPAGP-----CTDVIIREPP--NHTVTTEYWSQSFAT 630  
Db 979 VSIILASPTTE---QPTATVTIADSGGDDVQPGVTLVCSNPPCETHETGTT-----NTAT 1030  
Qy 631 TTTVT---GPPSG-----TDTV-----IIREPPNHTVTTEYWSQSY 664  
Db 1031 TTVANLGGHPQPTQVQVCDRQEAASLVSTVGQQNGSVVRVCSNPPCETHETGTT 1090  
Qy 665 ATTTT-----ITAPPGET-----DTVLIREPPNH-----T 689  
Db 1091 ATTATSNMAGHGCNPPCETHETGTTTATTAMSSVGANHORDARRACAAAGTPAVIRIS 1150  
Qy 690 VTTTEYWS-----QSYATTTVT-----AP-----PGETDVTVLIR 719  
Db 1151 VATGALEAAOGSKSOCQTRQTSATSTMTVMATGAPCSAGPGLLGPMSAREPGRSPAFVQ 1210  
Qy 720 EPP-----NHTVTTEYWSQSY-----ATT 740  
Db 1211 LAPLSSKVRLLSSPSIKDLPAGRSHAVSTAAMTRSSVAGCEPRMAPVCESLQGGSPSTTV 1270  
Qy 741 TVTA-----PPGTDVTIIREPPNHTVTTEYWSQSFATTTTVPAGGTDVIIEYMS 795  
Db 1271 TVTALEALLCPSATVQVCSNPPCET-----HETGTTTATTSNAGSAQVCSNPPCE 1323  
Qy 796 SSKISTSSNDITSII-----PSFSRPHVNVSTTSDLSLTFE-SSSMNTPTSISS 842  
Db 1324 THETGTTTATTTATNGTGQPEGGOOPAGRCETHQTTSTGCTMTMSVSGALLPDATSS 1383  
Qy 843 -----DCMLLSSTTLVTESETT-----TELICSDGKECSRLSSSGIVTNP--DSNES 888  
Db 1384 HRTVESLIEVAAAPSVTQACTALLAPFTQVCS-----NPPCEIHE- 1426  
Qy 889 SLVTSVTPASTMSDSLSTGISTATSSD-----NVKSGSVVTTSTSTVTIQTTPNPLS 943  
Db 1427 ---TCTHTTATVTSMNSQNDPPPAASQGEVESTQGDVNIITSSSAITT--TVSSLT 1481  
Qy 944 SEVSLTQLSSIPSVSESESKVFTSNGQNS--CTHDSQSTSTEI-----EIVTTSSTK 996  
Db 1482 RAVTTVTQSTVPVPGSPVPEELQVSPGRQOLPPRQLQASASTALMGSAEVLASQTP 1541  
Qy 997 VLPVPVSSNDLTSEPTNTREQPTTLSTNSITEDITTSQPTGDNDNTSTNPVPTVA 1056  
Db 1542 ELPAADVLLSS--TGEPPSSGQ-----SAGSAVATVAVVQ-----PPPTQS 1580  
Qy 1057 TSTLASASER---DNKSGSHESASTSLKPSMGNSGLTSTTEIATTTSTPEAPSAPVSS 1113  
Db 1581 EVDQLSLPOELMAEAGTTLTLMVTLGP---EELAVTAAAEAAQAAATEAQALQA 1637  
Qy 1114 GTDV-----TTEPTDREQPTTLSTTSKTNSELVATTOATNENGKSPSTDLTSLTT 1166  
Db 1638 VLQAAQAVMGTPGEMDMTSEAAATV-----TOAEL---GHLSAEQEQCATTIPIVLTO 1688  
Qy 1167 GTSASTSANSSELVTSQS-----VTGGAVASASNDQSHSTSVTNSNSIVSNTPQTT-- 1216  
Db 1689 QELAAALVQOQLQEAQAQOHHLLPTEALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746  
Qy 1217 LSQOQVTSPPSTNTFIA 1233  
Db 1747 LPSTATESLAPSNTEFA 1763

RESULT 10  
PCT-US93-11721-5  
; Sequence 5, Application PC/TUS9311721

GENERAL INFORMATION:  
APPLICANT: Lamarco, Kelly  
APPLICANT: Wilson, Angus  
APPLICANT: Herr, Winship  
TITLE OF INVENTION: A NOVEL EKARYOTIC TRANSCRIPTION PROTEIN:  
TITLE OF INVENTION: HOST CELL FACTOR  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FLEHR, HOUBACH, TEST, ALBRITTON & HERBERT  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-4187  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11721  
FILING DATE: 03-DEC-1993  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman, Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: PP-57503-1/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2035 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-11721-5

Query Match 6.2%; Score 404.5; DB 5; Length 2035;

Best Local Similarity 21.9%; Pred. No. 3.7e-15;

Matches 319; Conservative 165; Mismatches 506; Indels 467; Gaps 69;

QY 14 IASAKTITGVDFDSFNSNAANYAFKGPYPTNAV-LGWSLDGTSA-NPQDTFLNM 71  
DB 537 IGSSPQNSGN---AALAAAAATQKIPPSAPTVLSVPACTTIVKTMVTPG---TTTL 589  
72 PCVEKYTTQSVDLTADGVKYATCFYSGEFTFTSLCTVNDALKSIRAFGTPLP 131  
DB 590 PATVKVASSPVWV-----SNPATRLKTAQAQVGTSSV---ATKSTRP 631  
QY 132 IAFWGGTGSSTLDESKCFAGTNTVTFMDGDKOISIDVEFEKSTVDPDSAYLYASRVP 191  
DB 632 I-ITVHKSGLTVVAQAQVTVTVVGGVT-----KTTL-----VKSPISVPGGSAIIS 678  
QY 192 SLNKVTVLFVAPQENGVTSGTGMGFSNSGNDVAIDCSNIHIGITKGLNDMNPVPSSESFS 251  
DB 679 NLGKVMVW---QTRPVGTSAVTC-QASTGPVT-----QIIQTKG-----714  
QY 252 YTKTCTSGNIQIKYQNPAGYRPFIDAYSATDVNOYTLAYTNDYTCAGSRLOSQKPTLR 311  
DB 715 -----PLPAG---TILKLVTSADGKPTTIIITQASGAG-----TKPTTL- 751  
QY 312 WTGYKNSDAGSNGIVAVTRTVTDSTTAVTTLPFNPSVDK-----TKTIELAPIP-- 363  
DB 752 -----GISVSPTSSTKPTGTTTIKTIPMSAIITQAGATGVTSSPGIKSPITII 799  
QY 364 TTTITTSYGVVTTSLT-----KTAPIGETATVVDVPHVHTTTVT 404  
DB 800 TTKVMTSGTGAPAKIITAVPKIATGHGOQGVTVVVLKGAP-GQPGTILRTVPMGVRLVT 858

QY 405 SEMTGTITTTTTRNPTDSIDTVVVOVPLPNPTVSTTEYWSQSFATTTTTAPPGGTDTV 464  
DB 859 -----PVTVSAVKP--AVTTLVVKGTGTGVTILGVT-----CTVSTSLAGAGHST- 902  
QY 465 IIREPPNHTVTTTEYWSQSFATTTTTAPPGGTDSVLIIREPPNPTVTTTEYWSQSFATTTT 524  
DB 903 -----SASLATPITL-----GFIATLSQVINPTAITVS-----NAQT 936  
QY 525 TVTAPPGGTDSVLIIREP-PNPTVTTTEYWSQSFATTTTTAPPGGTDSVLIIREPPNHTV- 582  
DB 937 TLTAAAGLTPPTITMQPVSOPT-----QVTLITAFSG-----VEAQPVDLIP 978  
QY 583 -----TTTEYWSQSFATTTTTAPPG---GDTVLIIREPP--NHTVTTTEYWSQSFA 630  
DB 979 VSILASPTTE---OPTATVTIADSCQGDVQGTVTILVCSNPPCETHETGTT-----NTAT 1030  
QY 631 TTTVT---GPPSG-----TDTV-----IIREPPNPTVTTTEYWSQS 664  
DB 1031 TTVANLGGHPQPTQVQFVCDREAAASLVSTVTVGOONGSVVYVCSNPPCETHETGTTNT 1090  
QY 665 ATTTT-----ITAPPGET-----DTVLIIREPPNH-----T 689  
DB 1091 ATTATSNMAGHGCNPPCETHETGTTNTATTATAMSSVGANHQDARRACAAAGTPAVIRIS 1150  
QY 690 VTTTEYWS-----QSYATTTVT-----AP-----PGETDTVLIR 719  
DB 1151 VATGALEAAOGSKSQCTROTTSATSTTVMATGAPCSAGPLLGPSMAREPGGRSPAFVQ 1210  
QY 720 EPP-----NHTVTTTEYWSQS-----ATTT 740  
DB 1211 LAPSSKVRLLSPSINKDLPAGRHSHAVSTAAMTRSSVGAEPMAVPCESLQGGSPSTTV 1270  
QY 741 TVTA-----PPGGTDTVLIIREPPNPTVTTTEYWSQSFATTTTTAPPGGTDTVLIIESMS 795  
DB 1271 TVTAEALLCPSATVTQVCSNPPCET-----HETGTTNTATTNAGSAQVCSNPPCE 1323  
QY 796 SSKISTSSNDITSI-----PSFSRPHYVNSTTSDLSSTFE-SSSMNTPTSIS 842  
DB 1324 THETGTTHTATTATNSGTGQPEGGOQPPAGRPCETHOTHTGTGTTMSVSGALLPDATSS 1383  
QY 843 -----DGMLLSSTTTLVTESETT-----TELICSDGKCEKRLSSSGIVTNP--DSNES 888  
DB 1384 HRTVESGLEVAAPSVTPQAGTALLAPPPTQVCS-----NPPCETHE- 1426  
QY 889 SIVTSTVPTATMSDLSSTDCIGSATSSD-----NVSKGVSVTTESTSTVTTIOTTPNPLS 943  
DB 1427 --TCGTHATTATVTSNMSSNODPPPAASDQGEVESTQGDVSNITSSAITT--TVSSTLT 1481  
QY 944 SSVTSLTQSLIPSSESKVFTFTSNCNQS--GTHDSQSTSTEI-----EIVTTSSTK 996  
DB 1482 RAVTIVTQSTPVPGPSVPPPELOVSPCPOQLPPQLLOQSASTALMGESAEVLSASQTP 1541  
QY 997 VLPVWSSNDLTSEPTNTRQOPTTLSTTSNITEDITTSOPTGNGDNTSSNPNVPTVA 1056  
DB 1542 ELPAADVLUSS--TGEPSGQE-----SAGSAVAVTVVYVQ-----PPTQS 1580  
QY 1057 TSTLASAEE---DNKSGSHESASTSLKPSMGNSGLTTSSTIEATTSPTPEAPSPAVSS 1113  
DB 1581 EVDQLSLPQELMAEAQAGTTTILMVTGLTP---BELAVTAAAEAAAQAATTEEAQALAIQA 1637  
QY 1114 GTDV-----TTEPTDTRQOPTTLSTTSKTNSELVATTQATNENGKSPSTDLTSLTT 1166  
DB 1638 VLQAAQAQVMTGTEPEMDTSEAAATV-----TOAEL-----CHLSAECQEGQATTIPVL 1688  
QY 1167 GTSATSELSVELTSGS-----VTGGAVASASNDQSHSTSVTNSNSIVSNTPTQT-- 1216  
DB 1689 QELAAALVQOQLOQEAQAQOQHHLPTLEALAPADSLNDPAIESNCL--NELAGTVPSTVAL 1746  
QY 1217 LSQOQVTSSTSPSTNTFIA 1233  
DB 1747 LPSTATESLAPSNTEFA 1763





STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-783-774-2

Query Match 5.0%; Score 323; DB 3; Length 907;  
Best Local Similarity 22.9%; Pred. No. 6.1e-11;  
Matches 266; Conservative 128; Mismatches 390; Indels 378; Gaps 60;

QY 152 TAGTN-TVTENDGDKDISIDVEFEKSTVDPSAVLYASRVMPSLNKNVTLFVAPQCENGYT 210  
DB 43 TADNVNINFDVGGKKHOLDLDFGOLTPHTKAVYQPRGAFGGSSENATNLFLELLGAGEL 102  
QY 211 SGTMGFSSNGDAIDCSNIHIGITKGLNDWNPVSSSEFSYTKTCSNGIOIKYONVPA 270  
DB 103 ALTM-----RSKKLPINVTGEEQ---QVSLES-----VDVYFQDV-- 135

QY 271 GYRPFIDAYISATDVNOYTLAYTNDYTCAGSRLOK---PFTLRWTGYNKSDAGSNGIV 326  
DB 136 -----FCGMCHHAEMQNPVYLIPETVPYIKWDCNCSNITAV 173

QY 327 IVATNTRVVT-----DSTTAVTTLPPNSVOKTKTI---EILQPIB-----TTT 366  
DB 174 VRAOGLDVLPLSLPSAODSNFSVKTEMLGNEIDECIMEDGELSQVLPGDKKNITCS 233

QY 367 ITTSYGVVTSYLTATIGETATVVDVYPHTTIVTSEWGTGTTTTTRINPTDSIDT 426  
DB 234 GYESHV-PSGGILSTSPV---ATPIGTGYAYSLRLPR-----PVSRLGNNSI-- 280

QY 427 VVQVPLPNPTVSTTEYWSQSFATTTTAPPGGTDVLIIEPPNHTVTTEYWSQSFAT 486  
DB 281 LVPYSGNGPKASGGDICIQS-----NIVSDEIFA-----SQDMPT 317

QY 487 TTTVTPPG--GTDSV--IIEPPN-PTVTTEYWSQSFATTT-----TVPA-PPGTD 534  
DB 318 NTDITVGNATYISVPWTSSEANSVPVTFWAFWNPNTDEFCKKWTLSGTPSGCE 377

QY 535 SVIIEPPNPTVTTEYWSQYATTTT--VTAPPGGTDSVVIIEPPNHTVTTEYWSQY 592  
DB 378 NI-----SGAFASNRFTDIVSGLGT-----APKTLIIITR----- 408

QY 593 ATTTVTAPPGTDVVIIEPPNHTVTTEYWSQSFATTTVTGPPSGTDVVIIEPP-- 650  
DB 409 ATNATTT-----THKVIKAPESITTSPLNTGTGADPNTTGLPSSHTVPTNLAPAS 463

QY 651 -NPTVTTEYWSQYATTTT-----ITAPPGETDVLIREPPNHTVTTEYWSQYATTTT 705  
DB 464 TGPVTSTADVTSPTAGTSGASVPTSPSPMDNGTESKAPDMT-----SSTSP 512

DB 706 VTAPPGETDVLIREPPNHTVTTEYWSQYATTTTITVAP--PGGTDVVIIEPPNPTVT 764  
DB 513 VTP-----TPNAT-----SPTPAVTTPTPNATS-----PTPAVTT 543

QY 765 TEYWSQSFATTTVTPPGGTDVVIIESSMSKISTSSNDITSIPFSRRHYVNSTS 824  
DB 544 -----TPPNATSPTLG-----KTSPTSAVTTPTPNATS-----PTLGK----- 576

QY 825 DLSTFSSNMTPTSSDGMLLSTTLVTESETTELICSDGKCSRSLSSSGIVNPD 884  
DB 577 ---TSPTSATVTPNA-----TSPTL-----GK-----TSPTSATVTP 608

QY 885 SNESSIVTVPVTPASTMSDLSLSDGISATSDSNVSKGSVVTETSVTTIOTPNPLSS 944  
DB 609 PN-----ATGPTVCTSPQANAT-----NHTLGG---TSPTPVVTSQ--PKNATS 648

QY 945 SVTSLTQLSSIPSSESSEKVTFTSNGDNGSDHDSQSTSEIEIVTTSKVKLPPVYSS 1004  
DB 649 AVT-----TCOHNITSSS-----TSSMSLRP---SS 671

QY 1005 NTDLTSEPTNREOPTTLS--TTSNSITD-----ITTSOPTGONGDNSTNPVPTVATSTL 1060  
DB 672 N-----PETLSPSTSDNSTSHMPLTSAHTPG--GENITQVTPA--SISTHHV 715

RESULT 13  
PCT-US95-04611A-19  
Sequence 19 Application PC/TUS9504611A  
GENERAL INFORMATION:  
APPLICANT: Soarete, Richard and Jackman, Winthrop, T.  
TITLE OF INVENTION: Non Splicing Variants of gp350/220  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Cooley Godward Castro Huddleson & Tatum  
STREET: 5 Palo Alto Square  
CITY: Palo Alto  
STATE: California  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/04611A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/229,291  
FILING DATE: April 18, 1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Luann Cseri  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: AVIR-003/000S  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-843-5163  
TELEFAX: 415-857-0663  
TELEX: 380816 CooleyPA  
INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 907 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-04611A-19

Query Match 5.0%; Score 323; DB 5; Length 907;  
Best Local Similarity 22.9%; Pred. No. 6.1e-11;  
Matches 266; Conservative 128; Mismatches 390; Indels 378; Gaps 60;

QY 152 TAGTN-TVTENDGDKDISIDVEFEKSTVDPSAVLYASRVMPSLNKNVTLFVAPQCENGYT 210  
DB 43 TADNVNINFDVGGKKHOLDLDFGOLTPHTKAVYQPRGAFGGSSENATNLFLELLGAGEL 102  
QY 211 SGTMGFSSNGDAIDCSNIHIGITKGLNDWNPVSSSEFSYTKTCSNGIOIKYONVPA 270  
DB 103 ALTM-----RSKKLPINVTGEEQ---QVSLES-----VDVYFQDV-- 135

QY 271 GVRPFIDAYISADVNOYTLAYNDYTCAGSRLOSK-----PFTLRMTGKKNSDAGSNGIV 326  
Db 136 -----EGTMMCHAMQNPVYLLIPETVPIKWDNCSTNITAV 173  
QY 327 IVATRTVT-----DSTAVTTLFPNSVDKTKTI---EILQIP-----TTT 366  
Db 174 VRAOGLDVTLPISLTSQAQSNFSEVTEMLGNEIDECIMEDGEISQVLPDKNKFNITCS 233  
QY 367 ITTSYGVTTSYLTAPIGETATVADVPHVHTTTVTSEWGTITTTTTRTNPTDSIDT 426  
Db 234 GYSEHV--PSGGLTSTSPV---ATPIGTGYAYSLRLTPR-----PVSRLGNNSI-- 280  
QY 427 VVVOVPLPNPTVSTTEYWSQSFAATTTTTPAPGGTDVVIIRPPNHTVTTTEYWSQSEAT 486  
Db 281 LVFVYNGKPGASGDYCIQS-----NIVFSEIPA-----SODMPT 317  
QY 487 TTTVTAPPG--GTDSPV--IIRPPN--PTVTTTEYWSQSFAATTT-----TVTA--PPGGTD 534  
318 NTDTITVYVGNATYSVPWVTSSEANSPNVTATFANPNNTEDFKCKWTLTSGTSGCE 377  
QY 535 SVIIRPPNPTVTTTEYWSQSFAATTTT--VTAPPGGTDVVIIRPPNHTVTTTEYWSQSY 592  
Db 378 NI-----SCGAFASNRFTDITVSGLT-----APKTLIIRT----- 408  
QY 593 ATTTVTAPPGTDVVIIRPPNHTVTTTEYWSQSFAATTTTTPGPGTDVVIIRPP-- 650  
Db 409 ATNATTT-----THKVFSAKPESTTSTPLNTTGFADPNTTGLPSSTHVPTNLTAPAS 463  
QY 651 -NPVTTTEYWSQSFAATTT-----ITAPGETDVLIRPPNHTVTTTEYWSQSFAATTT 705  
Db 464 TGPVSTADVTSPGTAGTSCASPVTPSPWDNGTESKAPDMT-----SSTSP 512  
QY 706 VTAPGETDVLIRPPNHTVTTTEYWSQSFAATTTTTPAP--PGGTDVVIIRPPNPTVTT 764  
Db 513 VTP-----TPNAT-----SPTPAVTTTPNAT-----PTPAVTT 543  
QY 765 TEYWSQSFAATTTTTPAPGGTDVVIIRPPNHTVTTTEYWSQSFAATTTTTPGPGTDVVIIRPPNHTVTT 824  
Db 544 -----PTPNATSPILG-----KTSPTSAVTTTPNAT-----PTLKG----- 576  
QY 825 DLSTFESSHNTPTSISSDGLMLSSLTTLVTESETTTELICSDGKCEKSLSSSGIVTNDP 884  
Db 577 ---TSPTSATVTPNA-----TSPTL-----GK-----TSPTSATVTP 608  
QY 885 SNESSIVTTPATMSDLSLSDGTSATSDNVKSGVSVTTTSTTITQTPNPLSS 944  
Db 609 PN-----ATGPTVGETSPQANAT-----NHLGG---TSPTPVTSQ--PKNATS 648  
QY 945 SVTSLTQLSIPVSESESKVTFTSNGDNQSGTHDSQSTSTIEIIVTTSKVLPPVSS 1004  
Db 649 AVT-----TSPTL-----TCQHNITSS-----TSSMSLRP---SS 671  
QY 1005 NDTLSTSEPTNTRQPTTILS--TTSNITED-----ITTSQPTGDNGDNTSSTNPTVATSTL 1060  
Db 672 N-----PETLSPTSDNSTSHMLLTSAPHTG--GENITQVTPA--SITHHV 715  
QY 1061 ASASEEDNKGSHESASTSLKPMGNSGLTSTTEIETAT--TSPTAPSPAVSGGTDVTT 1119  
Db 716 STSSPEP-----RPGTTSOAGSPGNSSTSTKPCENVNKTGTPQONATSPQAPSGQKTAV 769  
QY 1120 EPTDTRQPTTILSTSTKNSSELVATTOATNENGCKSPSDDLSTLTTCTGASTASSELV 1179  
Db 770 -----PTVSTGKANS-----TT-----GCKH-----TTGCHGARTSTPTD 802  
QY 1180 TSG--SVTGGAVASANDOSHSTSVTNSINVTNPTQTLTSSQVTSSTNPTIATSYD 1238  
Db 803 YGGSITPRPRYNATYLPSTSKLPRWFTSPVTTAQATVPVPTSPQFESNL---- 859  
QY 1239 SCSTIIOHSTWGLCLTLLSLFI 1260  
Db 860 SMLVQWAS--LAVTLLULLLV 879

RESULT 14  
US-08-616-844-40  
; Sequence 40, Application US/08616844  
; Patent No. 5849578  
; GENERAL INFORMATION:  
; APPLICANT: FALB, DEAN A.  
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDWARDS  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/616,844  
; FILING DATE: 15-MAR-1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/599,654  
; FILING DATE: 09-FEB-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/485,573  
; FILING DATE: 07-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/386,844  
; FILING DATE: 10-FEB-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: CORUZZI, LAURA A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-053  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1481 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-616-844-40

Query Match 5.0%; Score 322; DB 2; Length 1481;  
Best Local Similarity 21.2%; Pred. No. 1.4e-10;  
Matches 246; Conservative 159; Mismatches 428; Indels 330; Gaps 51;

QY 301 SRLOSKPFTLRWTKYKN---SDAGSNGIVIVATT---RTVTDSTTAVTTLFPNPSVDKT 353  
Db 134 STVSSKEGVWQVTSCKSHAADAPENLLAFTADAGRSGSSRTNFTILPVGYSL--- 191  
QY 354 KTEILOPIPTTTITTSVGVVTS-----YLTKTAPIGETAT----- 390  
Db 192 -----IATALTQSGLNLAESLHLPSSSEDEFERIAAFQKSGTASEMGTERAMG 241  
QY 391 VIVDVPHYHTTTVTSEWT-----GTITTTTTRTNPTDSIDFVVVQVPLNPVSTT 441  
Db 242 LSEWTVHQEATTSANSPFLPALEMGEITPSRKRNSS-----GPDLS-- 286  
QY 442 EYWSQSFAATTTT-----VTAPPGGTDVVIIRPPNHTVTTTEYWSQSFAATTTTVP 496  
Db 287 --WLHFTYRANSSPLLLOLSSPSETEKL-----NNSTGLQSSSVSQTMTMHVATVF 335



Fri Oct 4 08:13:45 2002

Qy	650	PNTPTVT-----TEYWSQSYAYTTT-----TAPPGTDTVL-IREP-----	685
Db	489	SDTVOGSGHTALCDKSYSESSSTSSSELNLSAPRGSTLEDSEPRQALGOSSANA	548
Qy	686	-----PNHVTTTT-----EYWSQSYA-TTTTVTAPPGTDTVLIREPN-----H	734
Db	549	EDRTSGVPSLGTHTLATVNGERTLSRVLLTNTSMSTSGEAGSPAAMPQETECASLH	608
Qy	725	TVTTTEYWSQSYATTTVTVAPPGGTDVTVIREPNPTVTTEYWSQSFAATT-----	776
Db	609	VNVTDDMGVLNLSLAASGALGVAGISYQVGRGTAIBQRTSSDHTDHTYLSSTTKGERAL	668
Qy	777	-TVTAPPGGTDTVIYESMSSSKISTSSNDITGIIPSTSRPHYVNSTSDLTGFESS---	832
Db	669	LSITDNNSSSD--IVESSTSYIKLNSSH---EYSSFS---HAQTERGNISYDGEYAQ	720
Qy	833	-----SMNTPTSSDGMLLSTTLVTESETTEELICSDGKECRLSSSSSGIVNTP	883
Db	721	PSTESPVLHTNLSPTPTINNPNTSVVLDTDAEYFVSDSSSSSSSSSSSGPPLPLPS	780
Qy	884	DSNESSIVTSTVPTASTMSDLSSTDGISATSNDYKSGVSVTETSTVTITQTTPNPLS	943
Db	781	VSOHHLFESSLPTRASVHLKTSDSATPWSSSPSLPVSLTSTSA-----PLS	832
Qy	944	SSVTSLTQLSSIPSVSES-ESKVT-----FTSNGDNQSGTHDSQOSTEIEIVT-	991
Db	833	VSOTTLPOSSSTPVLPRARETVPVTFQSTMTSMFMLHSSQTADLKASQSTPHOEKVI	892
Qy	992	-----TSSTKV-----LPB-VYSSNTP--LTSPNTREQPTTLSTTSNSITED	1032
Db	893	SKPSLSVLPTSTKAVTNTNSPLPSTLSTESTQTLPATSTNLQMSPTFTTIT-----	946
Qy	1033	ITTSQPTGDNMTSSNPNPT--VATSTLASASEDNKSGHESASTSLKPSMGENSEGL	1090
Db	947	LKTSQPLMTTPTGLSSTASLVGTPAVOTTAG---KQSLSTHPEI---LVFPQISTEGGI	999
Qy	1091	TTSTE---TEATVTSPTAPSAVSSGTDVTEPTOTREQPTTLSTTSKTNSELVATTQA	1147
Db	1000	STERNRIVADATT-----GLIPTSVPTSAKDMTKLGVT-----AEYSP	1039
Qy	1148	TNENGKSKSPSTDLT-----SSLTGTSTASTANSSELVTSGVTCGVASASNDQS	1197
Db	1040	ASRLAGTSPSPOTTWSTAEIDLAPKATFAVQSSQSTPTTLSSSASVNSCAV-----	1091
Qy	1198	HSTSVTNSNSIVSNTPQ-----TTLSQQVTSSTSPSTNTPTIASTYDGSGS	1241
Db	1092	-NPLCHNGECVADNTSGYHCRPPSGWGDGCSVDVNECLSNPCPST---ATCNNTQGS	1146
Qy	1242	LIIOHSTWLY---GLITLLLSLFI	1260
Db	1147	FICKPVGVOLEKICNLVTFV	1169

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:29:03 ; Search time 29.15 Seconds  
(without alignments)  
4153.433 Million cell updates/sec

Title: US-09-715-876-8  
Perfect score: 6495  
Sequence: 1 MLOQFTLLFLYLSTASAKTI.....SIQHSWLYGLITLLSLFI 1260  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6495	100.0	1260	2 S60896	agglutinin-like pr
2	3194	49.2	1419	2 T30531	agglutinin-like ad
3	939	14.5	1367	2 S31939	hypothetical prote
4	739	14.5	1537	2 S33465	flocculation prote
5	711	10.9	1075	2 S48992	flocculation prote
6	679.5	10.4	1367	1 S48478	glucan 1,4-alpha-g
7	649.5	10.0	1275	2 T33369	hypothetical prote
8	646.5	10.0	2232	2 T34434	hypothetical prote
9	637.5	9.8	2271	2 F90073	hypothetical prote
10	632.5	9.7	3570	2 T45025	mucin MUC5B, trach
11	599	9.2	3020	2 A43932	mucin 2 precursor,
12	578	8.9	1283	2 T39174	hypothetical Serin
13	577	8.9	1161	2 S57180	probable membrane
14	553.5	8.5	1459	2 T32271	hypothetical prote
15	547.5	8.4	1609	2 S25345	probable membrane
16	526	8.1	1229	2 T25697	hypothetical prote
17	526	8.1	1630	2 A53577	ascites stialoglyco
18	515	7.9	4776	2 E95206	cell wall surface
19	511.5	7.9	3507	2 T34513	hypothetical prote
20	503.5	7.8	1777	2 T34369	hypothetical prote
21	501	7.7	1169	2 T38181	flocculation prote
22	492.5	7.6	1104	2 S59310	hypothetical prote
23	487	7.5	1832	2 T31113	probable membrane
24	482.5	7.4	1829	2 T24583	hypothetical prote
25	464.5	7.2	670	2 T36791	hypothetical prote
26	460	7.1	770	2 T22808	hypothetical prote
27	451.5	7.0	786	2 T16509	hypothetical prote
28	445	6.9	1802	2 S69703	HKR1 protein precu
29	443.5	6.8	1032	2 T34433	hypothetical prote

30	443.5	6.8	1251	2 T21389	hypothetical prote
31	435.5	6.7	2468	2 A83412	hypothetical prote
32	430.5	6.6	866	2 T45462	membrane glycoprot
33	428	6.6	867	2 T45463	membrane glycoprot
34	424	6.5	528	2 T47141	gastric mucin (clo
35	423.5	6.5	520	2 S62521	hypothetical prote
36	419	6.5	973	2 T40778	hypothetical prote
37	418.5	6.4	725	2 A41258	a-agglutinin core
38	415.5	6.4	1306	2 S25370	MSB2 protein - yea
39	412.5	6.4	3029	2 S76109	hypothetical prote
40	412	6.3	851	2 T22696	hypothetical prote
41	409.5	6.3	1041	2 S53862	probable membrane
42	405	6.2	833	2 E90577	lipoprotein vsac (
43	404.5	6.2	2035	2 A40718	host cell factor C
44	401.5	6.2	648	2 PC4395	mucin 3 - human (f
45	400.5	6.2	948	2 T11678	hypothetical prote

ALIGNMENTS

RESULT 1

S60896

agglutinin-like protein - yeast (Candida albicans)

C;Species: Candida albicans

C;Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 17-Mar-2000

C;Accession: S60896

R;Hoyer, L.L.; Scherer, S.; Shatzman, A.R.; Livi, G.P.

Mol. Microbiol. 15, 39-54, 1995

A;Title: Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual

A;Reference number: S60896; MUID:95272392

A;Accession: S60896

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1260 <HOY>

A;Cross-references: EMBL:L25902; NID:g704426; PIDN:AAC41649.1; PID:g704427

C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 100.0%; Score 6495; DB 2; Length 1260;

Best Local Similarity 100.0%; Pred No 1,7e-273; Mismatches 0; Indels 0; Gaps 0;

Matches 1260; Conservative 0;

QY	1	MLOQFTLLFLYLSTASAKTI	IGVDFDSFSLTWSNANYAFKGGYPTWNAVLSLDGTS	60
DB	1	MLOQFTLLFLYLSTASAKTI	IGVDFDSFSLTWSNANYAFKGGYPTWNAVLSLDGTS	60
QY	61	ANPGDTFTLMPCEVKYTT	SQTSVDLTADGVKATCOFYSGEEFTFTLCVTNDALKS	120
DB	61	ANPGDTFTLMPCEVKYTT	SQTSVDLTADGVKATCOFYSGEEFTFTLCVTNDALKS	120
QY	121	SIKAFGVTPLPIAFNVG	TGSSDLEDSKCFCTAGTNTVTNFNDGDKDISIDVEFEKSTVDP	180
DB	121	SIKAFGVTPLPIAFNVG	TGSSDLEDSKCFCTAGTNTVTNFNDGDKDISIDVEFEKSTVDP	180
QY	181	SAYLYASRVNPSLNKVT	TLFVAPQCENGYTSCTMGFSSSNGDVAIDCSNIHIGITKGLND	240
DB	181	SAYLYASRVNPSLNKVT	TLFVAPQCENGYTSCTMGFSSSNGDVAIDCSNIHIGITKGLND	240
QY	241	WNYPVSSSFYTKTCT	NSNGIQKYONVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG	300
DB	241	WNYPVSSSFYTKTCT	NSNGIQKYONVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG	300
QY	301	SRLOSKRFTLRWTG	VKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVDKTKTEILQ	360
DB	301	SRLOSKRFTLRWTG	VKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVDKTKTEILQ	360
QY	361	PIPTTTITTSYGVV	TSYVLTAKTAPIGETATVVDVPHHTTTVTSEMTGTTTTTTRNP	420
DB	361	PIPTTTITTSYGVV	TSYVLTAKTAPIGETATVVDVPHHTTTVTSEMTGTTTTTTRNP	420
QY	421	TDSIDIVVQVPLNP	IVPNTVTEWSQSFAITTTTAPGCTDVIIRPPPHVTTTEYW	480
DB	421	TDSIDIVVQVPLNP	IVPNTVTEWSQSFAITTTTAPGCTDVIIRPPPHVTTTEYW	480

```
Db 421 TDSIDTVVQVPLNPNTVSTTEYWSQSFATTTTATPGGTDVLIIRPPNHTVTTTEY 480
Qy 481 SOSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIR 540
Db 481 SOSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIR 540
Qy 541 PPNPTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTT 600
Db 541 PPNPTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTT 600
Qy 601 PPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEY 660
Db 601 PPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEY 660
Qy 661 SOSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIR 720
Db 661 SOSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIR 720
Qy 721 PPNHTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTT 780
Db 721 PPNHTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTT 780
Qy 781 PPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEY 840
Db 781 PPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEY 840
Qy 841 SSDGMLLSSTLVTSEETTELICSDGKCSRLSSSSGIVNPDNNESSIVTSTVPTAST 900
Db 841 SSDGMLLSSTLVTSEETTELICSDGKCSRLSSSSGIVNPDNNESSIVTSTVPTAST 900
Qy 901 MSDSLSDGIGSATSDNVKSGVSVTTSTVTTTQTPNPLSSSVSLTQLSSIPSVSE 960
Db 901 MSDSLSDGIGSATSDNVKSGVSVTTSTVTTTQTPNPLSSSVSLTQLSSIPSVSE 960
Qy 961 SESKVTFTSGNDQSGTHDSQSTSEIEIVTTSSTKVLPPVVVSSNTDLTSEPTNTREQPT 1020
Db 961 SESKVTFTSGNDQSGTHDSQSTSEIEIVTTSSTKVLPPVVVSSNTDLTSEPTNTREQPT 1020
Qy 1021 TLTSTTSNITEDITTSQPTGNDGNTSSTNPVPTVATSTLASAEDNKGSGHESASTSL 1080
Db 1021 TLTSTTSNITEDITTSQPTGNDGNTSSTNPVPTVATSTLASAEDNKGSGHESASTSL 1080
Qy 1081 KPSMGNSGLTSTTEATTTSPTEAPSPAVSSGTDVTEPTDREOPTTLSTSKTNSE 1140
Db 1081 KPSMGNSGLTSTTEATTTSPTEAPSPAVSSGTDVTEPTDREOPTTLSTSKTNSE 1140
Qy 1141 LVATTOATNKGKSPSTDLTSLTGTSASTANSELVTSSTGCAVASANDQSHST 1200
Db 1141 LVATTOATNKGKSPSTDLTSLTGTSASTANSELVTSSTGCAVASANDQSHST 1200
Qy 1201 SVTNSINVTNPOTTLTQQVTSSTPSTNFTASTYDGSGLIOHSTWLYGLITLLSLFI 1260
Db 1201 SVTNSINVTNPOTTLTQQVTSSTPSTNFTASTYDGSGLIOHSTWLYGLITLLSLFI 1260

RESULT 2
T30531
agglutinin-like adhesin - yeast (Candida albicans)
C:Species: Candida albicans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000
C:Accession: T30531
R:Gaur, N.K.; Klotz, S.A.
Infect. Immun. 65, 5289-5294, 1997
A:Title: Expression, cloning, and characterization of a Candida albicans gene, ALA1, tha
A:Reference number: Z20847; MUID:98053977
A:Accession: T30531
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1419 <GAU>
A:Cross-references: EMBL:AF025429; NID:g2522218; PID:g2522219; PIDN:AAB88883.1
C:Genetics:
A:Gene: ALA1
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
```

```
Query Match 49.2% Score 3194; DB 2: Length 1419;
Best Local Similarity 49.5% Pred No 9, 8e-131;
Matches 708; Conservative 182; Mismatches 355; Indels 186; Gaps 29;

Qy 1 MLOQFTLLFLYLSTASAKITGTFDSFNSLWSSMAANYAFKPGPYTNVAVLWGLSDGTS 60
Db 1 MLOQFTLLFLFLFAKATGTFNSIDSLWSSNAGNYAFKPGPYTNVAVLWGLSDGTS 60
Qy 61 ANPGDTFLAMPVCFKTYTSQTSVDLTADGVKATCOFYSGBEFTTSTLTCTVNDALKS 120
Db 61 ANPGDTFLAMPVCFKTYTSQTSVDLTADGVKATCOFYSGBEFTTSTLTCTVNDALKS 120
Qy 121 SKAPGVTVLPANFNGGTSSTDLSDKCFCTAGTNTVTFNDGSKKLSIAVNEKSTVDP 180
Db 121 SKALGVTVLPANFNGGTSSTDLSDKCFCTAGTNTVTFNDGSKKLSIAVNEKSTVDP 180
Qy 181 SAYLVSVMPSLKNKVTTLFVAPQCENGYTSGTMGSSNGDVAIDCSNTHIGTKGLND 240
Db 181 SGYLTTSRFMPSLKNKATLYVAPQCENGYTSGTMGSSNGDVAIDCSNTHIGTKGLND 240
Qy 241 WNPVSSSESYTKTCTNSGIOKYNVPAGYRPFIDAYISATDVNOYTLAYTNDYTCAG 300
Db 241 WNPVTSSESYTKSCSSEGISITQYNVPAGYRPFIDAYISPDNNQYQLSYKNDYTCVD 300
Qy 301 SRLQSKPFTLRWTKYKNSDAGSNGIIVVATRTVTDSTTAVTLPFPNSVDKTKTIEILQ 360
Db 301 DYWOHAPFTLRWTKYKNSDAGSNGIIVVATRTVTDSTTAVTLPFPNSVDKTKTIEILQ 360
Qy 361 PIPTTTITTSVGVVTSYLTKTAPIGETATVIVDVVPYHTTTVTSEMTGTTTTTTRNP 420
Db 361 PIPTTTITTSVGVVTSYLTKTAPIGETATVIVDVVPYHTTTVTSEMTGTTTTTTRNP 420
Qy 421 TDSIDTVVQVPLNPNTVSTTEYWSQSFATTTTATPGGTDVLIIRPPNHTVTTTEY 480
Db 421 TDSIDTVVQVPSNPNTTITQFNSSEPTSTTTITNSLKGTDVSVIVREPHNPVTTTEF 480
Qy 481 SOSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIR 540
Db 481 SESFATTTITTSKPECTDSVIVREPHNPVTTTEFWSSEYATTTIINGPEGSDVIVRE 540
Qy 541 PPNPTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTEYWSQSFATTTT 600
Db 541 PPNPTVTTTKFNSSEYATTTITNKPECTDSVIVKEPNPTVTTTEFWSSEYATTT 600
Qy 601 PPGGTDVLIIRPPNPTVTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTT 657
Db 601 GPEGTDVLIIRPPNPTVTTTEFWSSEYATTTIINGPEGSDVIVRE 660
Qy 658 -----EYWSQSFATTTTATPGGTDVLIIRPPNHTV----- 691
Db 661 SDDSNISSAQSSSESVQSSIVGLSSSDIPLSSDMPSSSGLTSSSESVSYSDSD 720
Qy 692 -----TTEYWSQSFATTTTATPGGTDVLIIRPPNHTV----- 725
Db 721 SSSSIESTLSSDRCSISIDTTFWSSSSDLESTITWSSSIDAQSSHLVQSVNSI 780
Qy 726 VTTTEYWSQSFATTTTATPGGTDVLIIRPPNPTVTTTATPGGTDVLIIRPPNPTVTT 768
Db 781 STSOLSSSSSEESTF-----ATDALVSSDASILLSSDSSSPSSDQSDILTSSSPS 835
Qy 769 -----SOSFATTT-----TVTAPPGT-DTVIYESMSSSKISTSSNDI----- 806
Db 836 GESDLSISITFTVEISSDSVSLTSDPASFSDDLSSSLNSDSSSPSSDQSDILTSSSPS 895
Qy 807 TSIIPSPS-----RPHYVNSTSLSTFESSMNTPTSISSDGMILLSTTLVTPES 856
Db 896 TLVVPSPFLSSSSSLUTPHYVNSTTVHASSESSSVASPSMASAN--DDTYTLSES 953
Qy 857 EPTTELICSDGKE---CSRLSSSGIVN-PDNSESVITSVPTASTMDSLSSTDG-I 911
Db 954 TDTTSSIGTSSVTFCRDNGDGCIVTGMPESSSIDSEQTSDVTTTSSFVASSTPTSAEQ 1013
```



QY	912	SATSSDNV--SKSGVSVYTTSTSIQTTPNPLSSVTSITOLSSIPSVSESKVTTTS	969
DB	1014	SITDNPNDSSQTSASSSTKSSVSVDRVVNSILLSETS--TLSSDDSTN--SDTSISSTT	1070
QY	970	NGD---NQSCTHDSQSTSEIEIVTTS--SKVLPPVVSNTOLSEPTNTR-----	1016
DB	1071	NSDTGMINAGSHSTASIKESIOKTCVILSSYLSTKLSTSDITLIELITLITELTEL	1130
QY	1017	-----EOPPTLSTNSIIE-----DITTSQTPGNGDNTSSNPNPVAT-----	1061
DB	1131	TTTEDNEPNTFTSPSHSEITSSONSVLKQVDRSTIKTSPPTDVTTVSSLSVHSTEA	1190
QY	1062	SA-----SEEDNKGSHESASLKLPSKMGNSGLTGTSTIEATTTSTPEAPSVSSGTD	1116
DB	1191	STATLGNSFNSVASTPSNIATYLSLSTSSSNHATSSGTVKSEASAAEATPSPTTS--	1248
QY	1117	VTEPTDTRQOPTTLSTTKTN-----SELVATTOATN---ENGKGSFSDTSLTSTTGT	1168
DB	1249	NRLSYSTEARAGITYANGSGSTNLNLTESQVAAPTDSTSVLIENPVVTTSTPDONSAAV	1308
QY	1169	SASTSANSLEY---TSGSVTGCAVASNDQSHSTVS-----TNSNI	1208
DB	1309	PSKTSKIEESTMNPDSNTNNGFTATLSQAQVPSSSIHSELITTTAKTTDASMGDSA	1368
QY	1209	VSNTPTTLQQVTSSTSPSTNTFIATYDYGSGIIHQHSTWLYGLITLLSLF	1259
DB	1369	ASNSQPTTLIOQVATS--SYNQLTITTVAGSSATKHPSSLKLFISVALFF	1417

RESULT 3

S51959  
 hypothetical protein YAL063c - Yeast (Saccharomyces cerevisiae)  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 05-May-1995  
 C:Accession: S51959  
 R:Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Ouel  
 submitted to the EMBL Data Library, August 1994  
 A:Description: The sequence of chromosome I of Saccharomyces cerevisiae.  
 A:Reference number: S51956  
 A:Accession: S51959  
 A:Molecule type: DNA  
 A:Residues: 1367 <BUS>  
 A:Cross-references: EMBL:U12980; GSPDB:GN00001; MIPS:YAL063c  
 C:Genetics:  
 A:Gene: MIPS:YAL063c  
 A:Map position: 1L

Query Match 14.5%; Score 939; DB 2; Length 1367;  
 Best Local Similarity 28.1%; Pred. No. 2.9e-33;  
 Matches 410; Conservative 214; Mismatches 469; Indels 364; Gaps 84;

QY	28	NSLTWSNAANYAFKPGPGYTWNAVIGLSWLSQCTSANPGDFTFLNMPCKVFKYTTISQT---	83
DB	50	DSSTYSNAYWAY---CYASKT-----KLASVGQGTDSIDYNIPCV---SSSGTFPCPQ	98
QY	84	VDLTAD-GVK-YATQC-----FYSGEFTFTFLCTVNDALKSSIKAFGWVLPPIA--	133
DB	99	EDLYGNWCKGIGACSNPIAYWSTDLFGFYTPT-----NVTLEMTGY	143
QY	134	FNVGGTGSGTLEDSKCFCTAGTNTVTENDGDKDISIDVEFEKSTVDPSAVLYASRVMSL	193
DB	144	FLPPTGSGYT-----FKFAVDDSDAIL-----	165
QY	194	NKVTTFLVAPOCNGYTGTHGTFSSNCDVAIDC-----SNHIGITKGLNDWN-	242
DB	166	-----SVGSEIAFCCAQEQPITSTNTI---NGIKPWNQ	198
QY	243	YPVSSSEFSYTKTCTGNGIQIKYQNVPA--GYRPF-----IDAYSATDYN-	286
DB	199	SPDNTITGVNVMYAGFYYPKMIYVNAVAMGTLPISVLPDGTTVSDDFEGYVTFDNNL	258

```

Db 1290 VIGHSSVSVSTGNTKSLTSLGSLTMSQOPRSTPASSVMGVSSTASLEISTYAGSANSLS 1349
QY 1244 QHSTWLYGLTLLSLFI 1260
Db 1350 LAGSGLSVFTASLLALAI 1366

RESULT 4
F1000 protein FLO1 precursor - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein FAR050W
C:Species: Saccharomyces cerevisiae
C:Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 29-Oct-1999
C:Accession: S53463; S7851; S31230; S57349
R:Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Kabac
submitted to the EMBL Data Library, February 1994
A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of the S
A:Reference number: S53458
A:Accession: S53465
A:Molecule type: DNA
A:Residues: 1-1537 <BUS>
A:Cross-references: EMBL:L28920; NID:g1616966; PIDN:AAC09499.1; PID:g694125; MIPS:YAR050
R:Watarai, J.; Takata, Y.; Ogawa, M.; Sahara, H.; Koshino, S.; Onneda, M.L.; Airaksinen,
Yeast 10, 211-225, 1994
A:Title: Molecular cloning and analysis of the yeast flocculation gene FLO1.
A:Reference number: S43543; MUID:94262325
A:Accession: S43543
A:Molecule type: DNA
A:Residues: 1-428, 'M', 430-473, 'M', 475-518, 'M', 520-549, 'T', 551-608, 'L', 610-638-69
A:Cross-references: EMBL:X78160
R:Onneda, M.
submitted to the EMBL Data Library, September 1994
A:Reference number: S57851
A:Accession: S57851
A:Molecule type: DNA
A:Residues: 1-428, 'M', 430-463, 'D', 465-473, 'M', 475-518, 'M', 520-549, 'T', 551-608, 'L', 610-63
A:Cross-references: EMBL:X78160; NID:g535933; PIDN:CAA55024.1; PID:g535934
R:Teunissen, A.W.R.H.; Holub, E.; van der Hucht, J.; van den Berg, J.A.; Steensma, H.Y.
Yeast 9, 423-427, 1993
A:Title: Sequence of the open reading frame of the FLO1 gene from Saccharomyces cerevisi
A:Reference number: S31230; MUID:93289821
A:Accession: S31230
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-296, 927-1516, 'TAYWPVVV' <TEU>
R:Bidard, F.; Body, M.; Blondin, B.; Deguin, S.; Barre, P.
Yeast 11, 809-822, 1995
A:Title: The Saccharomyces cerevisiae FLO1 flocculation gene encodes for a cell surface
A:Reference number: S57349; MUID:96090130
A:Accession: S57349
A:Molecule type: DNA
A:Residues: 1243-1274; 1308-1339; 1359-1390 <RID>
C:Genetics:
A:Gene: SGD:FLO1
A:Cross-references: SGD:S0000084; MIPS:YAR050w
A:Map position: 1R
C:Keywords: duplication; glycoprotein; transmembrane protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-1537/Product: flocculation protein FLO1 #status predicted <MAT>
F:953-997/Domain: repeat A2 <RA2>
F:998-1042/Domain: repeat A3 <RA3>
F:1043-1081/Domain: repeat A4 <RA4>
F:1082-1276/Domain: repeat B1 <RB1>
F:1277-1284/Domain: repeat B2 (partial) #status atypical <RB2>
F:1291-1341/Domain: repeat B3 <RB3>
F:1342-1392/Domain: repeat B4 <RB4>
F:1408-1416/Domain: repeat C1 <RC1>
F:1417-1425/Domain: repeat C2 <RC2>
F:1426-1434/Domain: repeat C3 <RC3>
F:135, 187, 262, 1114/Binding site: carbohydrate (Asn) (covalent) #status predicted

```

```

Query Match 14.5%; Score 939; DB 2; Length 1537;
Rest Local Similarity 25.8%; Pred No. 3.3e-33;
Matches 414; Conservative 246; Mismatches 520; Indels 426; Gaps 84;

QY 2 LOQFTLLFLYLSTASAKT-----ITGVDFSE-----NSLTWSNAANYAF----- 40
Db 10 LAVFTLLAL-TSVASGATEACLPAGQKSKMNFYQYSLKDSSTYSNNAYMAYGYASKT 68
QY 41 -----KPGGYPTW-----NAVGLHSLD--G 58
Db 69 KLGSSVGOTDISIDYINPCVSSSGTFFPCQEDSYGNWCKGMGACNSQGIAYWSTDLFG 128
QY 59 TSANPGDTFLNM-----PCVKYITTSQ-----TSVD 85
Db 129 FYTTPNV-LEMTGFLPQPGSTYFKAATVDDSAILSVGGATAFNCAQOQPITSTN 187
QY 86 LPAQGVK-----YATCOFYSGEFTFTSLCTVNDALKSKIKAGT-----VTLP 131
Db 188 FTIDIKPWGSLPPNIEGTVMYAGYYPM-----KVYVSNAVSMGTLPISVYLP 238
QY 132 IAFNVGGTSGTDLDSKCFAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLYASRYM 190
Db 239 -----DGTTVSDDFEG-----YVYSFDD-----DLSQSNCTVPDPSNYA-VSTTT 277
QY 191 PSLNKVITLFAVQCENGYTGTMGSSSSNGDVAICDSNIHIGITKGLNDKNVPSSESF 250
Db 278 TTPTEPWGTGTSTSTMTTGTGNGVPTDETIVIRTPPTASTIIITTEPNWSTFTST 337
QY 251 SYTKTCTSNIGIQIKYONVAGYRPPIDAYISADVNQVTLAYNDYTCAGSRLOQSKPFTL 310
Db 338 ELTITVTGNGVRTDEII-----VIRTPPTATTAITTEPNWSTFTSTSTEL----- 384
QY 311 RWTGYKNSDAGSNGI---VIVATRTVTDSTTAVTTL-PFNPSVDKTKT----- 355
Db 385 -----TIVTGTNGLPDTETIIIVIRTPPTATTAITTEPNWSTFTSTSTELTIVTGTNGL 438
QY 356 -----TEILOPPTTTIT-----SYGVVTSVLTAKTA---PIGETATVIVDPVYH 398
Db 439 PTDETIIVIRTPPTATTAITTEPNWSTFTSTSTELTIVTGTNGLPDTET-IIIVIRPTT 497
QY 399 TTT-TVTSWMTGTII-----TTTTRN-PTDSIDTVVQVPLNPPTVSTT-----EY 443
Db 498 ATTAMTTQPNWDTFTSTSTETITVGTNGLPTD--ETIIV---IRPTTATTAMTTPOP 552
QY 444 NSQSFAAT---TTVTAPPG-GTD--TVIIRPPNHT---VTTEYWSQSFAAT---TTV 490
Db 553 NNDFTSTSTMTTGTGNGVPTDETIIIVIRTPPTATTAITTEPNWSTFTSTSTEMTV 612
QY 491 TAPPG-GTDS--VIRPPNPT--VTTEYWSQSFAAT---TTVTAPPG-GTDS--VII 538
Db 613 TGTNGLPDTETIIIVIRTPPTATTAITTEPNWDTFTSTSTETITVGTNGLPDTETIIVI 672
QY 539 REPPNPT--VTTEYWSQSFAAT---TTVTAPPG-GTDS--VIRPPNHT---VTTEY 587
Db 673 RTPTTATTAITTEPNWDTFTSTSTETITVGTGTPDTDETIIIVIRTPPTATTAMTTQP 732
QY 588 NSQSFAAT---TTVTAPPG-GTD--TVIIRPPNHTV--TTTEYWSQSFAAT---TTV 634
Db 733 NNDFTSTSTMTTGTGNGVPTDETIIIVIRTPPTSEGLISTTTEPWGTGTSTSTEMTV 792
QY 635 T---GPPSGTDTVIRPPNP--TVTTEYWSQSFAAT---TTITAP---PGETDTVLI 682
Db 793 TGTNGLPDTETIIIVIRTPPTSEGLVTTTTEPWGTGTSTSTETITVGTNGLPDTETIIVI 852
QY 683 REPPNHTV--TTTEYWSQSFAAT---TTVTAPPG-GTD--TVIIRPPNHTV--TTTEY 731
Db 853 RTPTSEGLISTTTEPWGTGTSTSTEMTTGTGNGVPTDETIIIVIRTPPTSEGLISTTTEP 912
QY 732 NSQSFAATTTVTAPPGTD-----TVIIRPPNPTV--TTTEYWSQSFAATTTVTAPP 782
Db 913 WTGFTSTSTEMHTVGTGNGVPTDETIIIVIRTPPTSEGLISTTTEPWGTGTSTSTETVTI 972
QY 783 GGT-----DTVIIYESMSSSKI-----STSSNDITSIIPIFSRP----- 816

```

```

Db 973 TCTNGQPTDEIVIRTPTEGLSTTTTPEWGTFTSTSTMTTGTNGQPTDETVIYI 1032
QY 817 -----HYNSTSD-LTFESSMNTPTSISSDGMLLSTTLVTESETT-----TELICS 865
Db 1033 RPTSEGLVTTTTEPMTGTFSTSTSTMTSTVGTNGLTDETVIWKTTPTTAISSLSSS 1092
QY 866 DGKCSRLSSSGIVTNP--DSNESSIVTSTVPTASTMSDLSLSDGISATSDSNVSKSG 923
Db 1093 SGQITSSITSRPIIT-PFYPNGTIVISSVITSSLTSSFTSPVSSVSSITSS 1151
QY 924 VSVTSTSTVTTIQTTPNPLSSSVTSLQSSIPSVSESEKVTFTSNGDNQSGHOSQST 983
Db 1152 TSIFSESSKSV--IPTSSSTSGSESETSSAGSVSSS----SFISSESSKSPYSSSS- 1204
QY 984 STEIEVTTSTST-----KVLPPVSSNTDITSEPTNTREOPTTLSTTS-----NSITE 1031
Db 1205 ---LPLVTSATTSEATSSLPATT-----TKTSEQTLTVTSCESHVCTESISP 1252
QY 1032 DITTSQPTGDNQNTSSNT--PVPVTATSLASAEEDNKGSHESASTSLKPSMGENGSG 1089
Db 1253 AIVSTATVTVSGVTVETTTMCPITSTETTKQGTTEQTTETTKQTTVVVITSSCEDVCS 1312
QY 1090 LTTSTTEATTSTPEAPSPAVSGDVTTEPTDREOPTTLSTT-----SKTNSLV 1142
Db 1313 KTASPAVSTSTATNGVTTEYTTMCPIST--TESROQTLTVITSCESGVCSSTASPAI 1370
QY 1143 ATT-----QATNKGKSPSTDLTSLTGP-SASTSANSSELVTSQSW 1184
Db 1371 VSTATVNDVTVVTPWRPOTANEESSVKMNSATGETTNTLAAETTNVAETIIN 1430
QY 1185 TCGA-----VASANDOSHSTV-----TNSNSTVNTPTTLTSLQO 1220
Db 1431 TGAATKTVVTSLSRSHAEQFASATDVIGHSSVVSVSETGNTKSLTSSGLSTMSQ 1490
QY 1221 VNSSP-----STWTFASVYDGSIIQHTWLYGLITLLSLFI 1260
Db 1491 PRSTPASSVGVSTASLEISTYAGSANSLLAGSLSVFIASLLAI 1536

RESULT 5
S48992
flucoculation protein homolog YHR211w - yeast (Saccharomyces cerevisiae)
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 29-Oct-1999
C:Accession: S48992
R:Macr1, C.
submitted to the EMBL Data Library, February 1994
Description: The sequence of S. cerevisiae cosmid 9177.
A:Reference number: S46671
A:Accession: S48992
A:Molecule type: DNA
A:Residues: 1-1075 <MAC>
A:Cross-references: EMBL:U00029; NID:g551322; PID:g458919; GSPDB:GN00008; MIPS:YHR211w
C:Genetics:
A:Gene: MIPS:YHR211w
A:Map position: 8R

Query Match 10.9%; Score 711; DB 2; Length 1075;
Best Local Similarity 25.9%; Pred. No. 1.5e-23;
Matches 324; Conservative 192; Mismatches 93; Indels 342; Gaps 66;

QY 153 ACTNTVTFDGDKDISDIVER-----EKSTVDPSAYL---YASRV----- 189
Db 23 SCATEACLPAGORKSOMNIFYQSLKSDSTYSNAYWAYGYASKTKLGSVGGQTDISID 82
QY 190 ---MPLANKVTLVAPOCENYTS-CTMGFSSNGDVAIDCSNIHIGITKGLNDWNPYS 246
Db 83 YNIPCVSSGT-FPCQ-EDSYGNWCKMGMA-----CSN-----SQIATW-----S 123
QY 247 SESPSYKTCSTNGIOIKYQNVNPGYRPFIDAYISATDVNQYTLATNDNYTCAGSR---L 303

```

```

Db 124 TDLFGFYVTPNTVLTLEMTGYFLPQGTGYTFSEATVDDSAILLSVSGSIAFECCAQEPPI 183
QY 304 QSKPFTLRMTGYKNSDAGSNGIVVAT-----RTVTDSTAVTTLTPNPSPVDKT 353
Db 184 TSTNFTI--NGKPNW--GSLPDNITGVTVMYAGYYPPLKVYSNAVSMGLTPI--SVE-- 236
QY 354 KTBILQIPT-TTITTSYGVVTSYLTAKTAPIGETATVIVDVPYHTT---TTVTSEWGT 409
Db 237 -----LPDGTIVSDNFEYVYSF---DDDLQSNCNCTIPDPSIHTTSTITTTTPWAG 285
QY 410 TIT-----TTTTRNPTDSIDTWVQVPLPNPVTSTTEYWSQSFATTTTVPAPGGTDT 463
Db 286 TFTSTSTEMTTITDNGQLTDERVIV---IRPT-----TASTIT----- 322
QY 464 VIIREPNHTVTTTEYWSQSFATT---TIVTAPPG-GTDS--VIIREPPNP--TVTTE 514
Db 323 -----TTTEPMTGFTSTSTSTMTTGTNGQPTDETVIIRTPTEGLITTTTE 371
QY 515 YWSQSFATT---TIVTAPPG-GTDS--VIIREPNP--TIVTTEYWSQSFATTTVAP 565
Db 372 PMTGTFTSTSTMTTGTNGQPTDETVIIRTPTEGLITTTTEPMTGFTSTSTSTVTT 431
QY 566 PGGDSVIIREPNHTVTTTEYWSQSFATTTTIVTAPPGTDTVIIREPNH--TVTTEY 623
Db 432 ITGNG---QPTDET-----IVIRTPTEGLITTTTEP 462
QY 624 NSQSFATT---TIVT---GPPSGTDTVIIREPNPTV--TTEYWSQSFATT---TTI 670
Db 463 WTGFTSTSTMTTGTNGQPTDETVIIRTPTEGLITTTTEPMTGFTSTSTSTVTTI 522
QY 671 TAPPG-TD--TVLIREPNH--TIVTTEYWSQSFATT---TIVTAPPG-TD--TVLI 718
Db 523 TGTNGQPTDETVIIRTPTEGLITTTTEPMTGFTSTSTMTTGTNGQPTDETVIIRTP 582
QY 719 REPNH--TIVTTEYWSQSFATTTTIVTAPPGT-----DTVIIREPNPTVTTTEYWSQ 770
Db 583 RPTSEGLITRTEPMTGFTSTSTSTMTTGTNGQPTDETVIIRTP--TAISSLS 641
QY 771 SPATTTVTA-----PPGQDVT---IYESMSSKISTSSNDITSIIPSFSRPHY 818
Db 642 SQGITSITSSRPIITPPYNGTSSVSSSVSSVSSVSSVSSVSSVSSVSSVSSVSS 695
QY 819 VNSTSDLSLTFESSMNT--PTSISSDGMLLSTTLVTESETTELICSDGKCSRLLSS 876
Db 696 ---STTSTISFSESSTVPIPTSSSTSGSEKTSASSSSSSSSSSSSSPKSPNSSSS 753
QY 877 SGIVTNPDSNSESIVTSTVPTASTMSDLSLSDGISATS-----SDNVKSGSVSVTET 930
Db 754 LPPVTSATQTQET--ASSLPATTTKTS-EQTLTVTSCESHVCTESISSAIVSTATVT 910
QY 931 ---SVTTIQTTPNPLSSSVTSLQSSIPSVSESEKVTFTSNGDNQSGHOSQSTSTEIE 988
Db 811 VSGVTVETTTMCPITSTET-----KOTKGTTEQTKGTTEQ 847
QY 989 IVTTSSTKVLPPVSSNTDLTSEPTNTREOPTTLSTTSNSITEDIITTSQPTGDNQNTSS 1048
Db 848 TETTKQTTVV--TSSCESDICKTAS---PAIVSTSTATI-----NGVTTEY 890
QY 1049 TNPVTATSLASAEEDNKGSHESASTSLKPSMGENGSLTSTTEIEATTTTPTPEAP- 1107
Db 891 TWCPISTTES-----KQQTLLVTVTSCESGVCSSTTSPA 925
QY 1108 ---SPAVSSGQDVT-----EPTDREOPTTLSTTSKTNSELVATTOATNKGKSPSTDL 1160
Db 926 IVSTATATVNDVTVVTPWRPQITNEQ-----SVSSKMNS---ATSETTNTGAETKAV 978
QY 1161 TSSLTGTSASTSANSSELVTSQSVTCGAVASANDQSHSTV-----TNSNIVSNTPQT 1215
Db 979 TSSLSRNRHAEQ-----TASATDVIGHSSSVSVSETGNTMSTSSGLS 1023
QY 1216 TLSQOVTSS-----SPNTNFTASYDGSIIQHTWLYGLITLLSLFI 1260
Db 1024 TMSQOPRSTPASSVSGSSTASLEISTYAGSANSLLAGSLSVFIASLLAI 1074

```

```

RESULT 6
S48478
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)
N:Alternate names: extracellular glucanase; mucin-like protein MUC1; protein YIR019c
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 12-Nov-1999
C:Accession: S48478; A26877; B26877; S27281; JG6123
R:Rowley, K.
submitted to the EMBL Data Library, October 1994
A:Reference number: S48478
A:Accession: S48478
A:Molecule type: DNA
A:Residues: 1-1367 <ROW>
A:Cross-references: GB:Z47047; EMBL:Z38061; NID:g603997; PID:g763364; GSPDB:GN00009; MIP
R:Kamashita, I.; Nakamura, M.; Fukui, S.
J. Bacteriol. 169, 2142-2149, 1987
A:Title: Gene fusion is a possible mechanism underlying the evolution of STAL.
A:Reference number: A91831; MUID:87194600
Accession: A26877
Molecule type: DNA
A:Residues: 1-242 <PAR>
A:Cross-references: EMBL:M16164; NID:gl72522; PIDN:AAA35014.1; PID:gl72525
A:Accession: B26877
A:Molecule type: DNA
A:Residues: 762-1331 <YAX2>
A:Cross-references: EMBL:M16165; NID:gl72523; PIDN:AAA35015.1; PID:gl72526
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.
FEBS Lett. 239, 179-184, 1988
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacchar
A:Reference number: S27281; MUID:89031230
A:Accession: S27281
A:Molecule type: DNA
A:Residues: 1-31 <PAR>
A:Cross-references: EMBL:X13857; NID:g4551; PIDN:CAA32069.1; PID:g4552
R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseudohy
A:Reference number: JG6123; MUID:96323237
A:Accession: JG6123
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1367 <LAW>
A:Cross-references: GB:U30626; NID:gl304386; PIDN:AAC49609.1; PID:gl304387
A:Gene: SGD:MUC1; DEX2; SGD:S0001458
C:Genetics:
Superfamily: MIPS:YIR019c; SGD:S0001458
Map position: 9R
Keywords: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase
F:5-21/Domain: transmembrane #status predicted <TM1>
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 10.48; Score 675.5; DB 1; Length 1367;
Best Local Similarity 26.28; Pred. No. 7,1e-22;
Matches 378; Conservative 177; Mismatches 535; Indels 353; Gaps 65;

Qy 1 MLOFTLLFLYLISAKTITGVDFNSLTSWNAANAFKGPQYFWNAVLGWSLDGTS 60
Db 1 MRPFLAYLVLSLL-----FNSAL-----GFPALVPRG-SSEGTSS 36
Qy 61 AN-----PGDTFTLNM--PCVFYTTSTQTSVDLTADGVKATQCFYSGEFTFTSLTC 112
Db 37 CNSLVNCPMLDFNMHQDQIMQYTLDTVTSVWQDNT-YQITIHVKGE----- 86
Qy 113 TVNDALKS---SIKAFGTVTLPIAFNVGGTSGSSDLDLSDKCF-TAGTWTVTNF-DGDKDI- 167
Db 87 --NDLKLSKLIIGVTKGPKGVOLGYNENTYLDNPTDFTA-----TFEYVATQDVN 139
Qy 168 -----SIDVEFEKSTVDPSAYLASRVNPLSKVITLFLVAPQCENGYTSGTMGFSSN 220

```



Query Match 10.08; Score 646.5; DB 2; Length 2232;  
 Best Local Similarity 23.54; Pred No. 2, 3e-20;  
 Matches 361; Conservative 227; Mismatches 543; Indels 405; Gaps 62;

QY 12 LSIAKATITGVFDS--FNSLTWSNAANAFKPGPYTNNAVLCWLDGTSANPGDTTL 69  
 DB 274 LSSSTIQSSGTFPSVASASTVSGTSAGSSATVSTIAGSTGATITPVGSSSTI 333  
 QY 70 NMPCVFYKTTSGTSDVLRADGVKATCOFYSGEETP-----STLCTVNDALKSSTKA 124  
 DB 334 G-----SSTPSASSSSG-----TMSITISGTSSTVTVVPGSSSTFASSTPIASSSSP 381  
 QY 125 FGTVTLPIAFNNGVT--GSSDPLEDKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 183  
 DB 382 GSTV--VAPGSSSTYGSSTPASAS-----SSSGTMTNSGTSSTVTV-----APVSSSTF 431  
 QY 184 LYASRVNPLSNKVTLLFVAPQCENGYTSGTMFSSNGDVAIDCSNIHIGITKGLDNWY 243  
 DB 432 GSSPTIASSSSGSGTSTVTVVSGSSSTYGSSTPASASAGTASTISG-----GTGSTATIV 486  
 QY 244 PVSSESP-SYTKTCTSNIGIQKYNVPAGYRPFIDAYISATDVNQYFLAVTNDYTCAGSR 302  
 DB 487 PGSSSSVSGSSTOSAPSS-----PCTMSTVSGTGTSTVTVVPGSSTPAPSS 533  
 QY 303 LQSKPFTLRWTKVKNDSAGNGIVI-----VATRTVTDSTAV-----TTLPEN 347  
 DB 534 SPNPSSSPASTGSTITISGSSSIIVSVGSGTVSGTGTSTLASSTATPGSSSTVPSS 593  
 QY 348 PSVDKTKTIELQIPTTTIT-----SYGVVTSYLTKTAP 384  
 DB 594 SSPQSSQ-----SPAPNGTSTTPSSQSSPSMNPSSSTPGSSQSTITPEGSTASP 649  
 QY 385 IGETATVIVDPVHTTNTVTSEMT-----GTITTTTTRNP-----TDSIDTV----- 427  
 DB 650 TGTGTSI-----FSVATEVTSQSTVPGSSSLGTQSTNSPSPSSLSLSTSGMSTLTSEPS 704  
 QY 428 -----VVQVPLENPVSTYEQW-SFANT-----TTVTAP----- 457  
 DB 705 PSTQSSGAQSLITTPSNPNQSSISLESTSGATITSSGACTITWSPSSQSSVSSGGS 764  
 QY 458 -----PG-----GTDVLIREPNNHTVTTTEYWSQSAT 486  
 DB 765 TSPAASTTSGEMTSQGTQTPGSSVSTSAALITSTQSVSTNSPGSTVTRPSTVSGSTSS 824  
 QY 487 TTTVTAPPGTD-----SVIIREP-----PNPT-----VTTTEYWSQSFAITTTTV 526  
 DB 825 GSTVTV--GSTEASTGSSVASSPAPSTQSNPNPSTSGSSMITQSPYPSOSTSPVSS 882  
 QY 527 TAPPGTDSVIIREPNPTVTTTEYWSQSVATTTTATPPGTDSDVIIREPNPHRTTTE 586  
 DB 883 TTPSPGS-----ATTTVTAPPGTDTVIIREPNHTVTTTEYWSQSFAITTTTVTPPSP 641  
 QY 587 YWSQSY-----PGTTLST---SPSPSQSTTIGTQGSTS-----PGISTTSEE 924  
 DB 925 MTSQSGTQTPGSGTGTVPSTVSDST--SSGSTVTVGSGSSSPIPSTQNTNPSTSS 982  
 QY 642 DTVIIREPNPTVTTTEYWSQSVATTTTATPPGTDVLIREP-PNHTVTTTEYWSQSY 700  
 DB 983 GSSMSTQTPQSSQSTSPVESSTGATSS--SGSPGTTLTISPSPSSTIGS-----SQGS 1037  
 QY 701 ATTTVTAPPGTDTVLIREPNNHTVTT---TEYWSQSVATTTV---TAPPGTD 750  
 DB 1038 TSPVSTISOGSTEI-----PGSGTGTVPSTVSGSAGSTATMGSTEAASSISGGS 1091  
 QY 751 TVIIREPNPTVTTTEYWS-----QSFATTTTATPPGTDVLIIEYMSKISITSS 803  
 DB 1092 T-----SPNPSSSTSPSTGATSPGSSGTTLTISPSPSOSTISGSSQSGSTSPVVSIT 1146  
 QY 804 NDITSIIPSPRHVNSTTSDLTSTESSMWTPTISISSDMLLSSTLTVTSETTTELI 863  
 DB 1147 GDMTS-OGSTQIPGSGTGTVPST--GSGSTISGSEITSQG-----STQTPRSLSTSPAI 1200

QY 864 CSDKECSRLLSSGIVTNPDSNESSIVT-STVPTAST-----MSDSLSTDGLSA 913  
 DB 1201 STSQSQSVNSPGSVITOPSVRCSTSSGSGTWTGTEGSGSTSSSNTSLSSSPVPS 1260  
 QY 914 TS-SDNVKSGSVYTE-----TSVTIQ-----TTPNPPLSSVTSLTQLSSIP 956  
 DB 1261 TSQSNPSTSGSSTPFPNPQSTSPVSTTIGEMTSHGSTQTPSTIGSTV---TQSTVS 1317  
 QY 957 SVSSEKSVTFTNGDNQSGTH-----DSQSTSTEIEIVTTS 993  
 DB 1318 GNSGSGSTVITGSEASTSGSFKTSPSSISPVTSPSPITPFASSTSGSTISDVSSVS 1377  
 QY 994 STKVLP-----PVSSNDLTSEPTNTREOPTTLT 1024  
 DB 1378 TTSLAPLSSSLPSTVPSSTQSFSTSECKASKSSPVQSVSTPNTPTGTESSSTLLSS 1437  
 QY 1025 TNSNTEIDITTSQPTGNDGNDTSTNPVPTATSLASASEEDNKSGHESASTSLKPSM 1084  
 DB 1438 TISGTOHTTMSK--ASSGTSPTNS--QTGSTVTMGSSS---TSGVSTSSASTQPM 1490  
 QY 1085 GENSGLTSTEIEATTSTPEAPSPVSSGTDVTEP---TDTREOPTTLSTTKTNE 1140  
 DB 1491 STSOGSAGSTVASSTAPASSTAPSTGTMSTSSGTVGSGTISESSTTASASQST 1550  
 QY 1141 L-----VATTOATN-----ENGGKSPSTDLTSLTGTASANSSELVTSGSV 1184  
 DB 1551 VTMGSSSTSGVSTSSASTOPOMSTSGSAGST--VASSTAGLVSTSTVPSSTGTMGST 1608  
 QY 1185 TGGAV-----ASASNDQ-----SHSTSVTNSNSIVSNTPTTSLQ----- 1219  
 DB 1609 SSGTVGTSISTSTASASQSGTVMGSSSTSGVSTSSASTOPOMSTSGSAGSTV 1668  
 QY 1220 -----QVTSSTSPNTFIASVDGS-GSTIOHST 1247  
 DB 1669 ASSTTGLVSTSTVPSSTGTMGSTSGCTVGSTISESS 1704

RESULT 9  
 F90073  
 Hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: F90073  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O-  
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: F90073  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2271 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13702612; PIDN:BA043752.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA2447

Query Match 9.8%; Score 637.5; DB 2; Length 2271;  
 Best Local Similarity 22.3%; Pred. No. 5, 9e-20;  
 Matches 330; Conservative 280; Mismatches 521; Indels 347; Gaps 52;

QY 17 AKTITGVDFNSLTWSNAANYAFKPGCYPTWNAVGLWLDGTSANPGDTFLNMPGVFK 76  
 DB 297 ATITGTRIDSNKSFHSGKVLGNKNGHGGNGDGGIGFAF-----SPG----- 339  
 QY 77 YTTQTSVDLTADGVKATCOFYSGEFTTFTLTCTVNDALKSSIK-----AFGTV 128  
 DB 340 -VLGETGNAAGVIGGLGNAP--GFKLDTHN-TSKPNSAANADPSNVAGGAFGAF 395  
 QY 129 TLPTAFNVGSGTSDTLED--SKETACTNTVTFNDGDKDISIDVEFEKSTVDPSPAYLYA 186







Matches 285; Conservative 120; Mismatches 438; Indels 262; Gaps 44;

```
QY 249 SFSYTKTCTKNGIQIKYQNV-----PAGYRPFIDAYISATDVNOYITLAYTNDYTCAGSR 302
Db 1236 AFCWYICGNGCTVKEHFNIGSITITRPSLTFTTITLPTPTS-FTTTTTTTPISSTV 1294
QY 303 LOSKP-FTLWMTGYKNSDAGSNGI-----VIVA-----TTRVTVDSTAVT----- 342
Db 1295 LSTTPKLCCLSDWLNEDHPSGSDGDRPPDVGCGAPEDIECRVADPHLSLEQHGOK 1354
QY 343 -----TLFPNPSVDKTKIE-----ILOPIPTTTITTSYGVGT 376
Db 1355 VOCDSVGFCKNEQFGNGFGLCYKIRVNCWPMKCIITPSPTTTPSPPTTTT 1414
QY 377 SYLTKTAPIGETATVIVDPVHTTTVTSEWGTITTTT---RTNPDSIDVWVQVPL 433
Db 1415 TLPTTTPSPPTT-----TTPPTTTPSPPTTITTLPTTTPSPPTTTPPTT 1467
QY 434 PNPVTST---EYWSQSFATTTVTAPGGDVIIIR---EPPNHTVTTEYWSQSFAT 486
Db 1468 TTPSPPTTTPSPPTTTPSPPTTTPPTTTPSPPTTTPPTTTPPTTTPPTT 1527
QY 487 TTTVTAPPGTDSVIREPPNPTVTTEYWSQSFATTTVTAPPGTDSVIREPPNPTV 546
Db 1528 TTTTTPPTTTPS-----PPTTPTTPTTSTTLPTTTPSPPTTTPPTTTPPTT 1582
QY 547 TTTTEYWSQYATTTVTAPPGTDSVIREPPNPTVTTEYWSQYATTTVTAPPGTD 606
Db 1583 TTT-----PSPPTITTTTTPPTT-----PPTTTPPTTTPPTTTPPTTTPPTT 1633
QY 607 TTIIREPPNHTVTTEYWSQSFATTTTTPGSDVIREPPNPTVTTEYWSQYAT 666
Db 1634 TL-----PPT-----PSPPTTTPPTT-----PSPPTTTPPTT-----PSPPT 1669
QY 667 TTTITAPPGTDSVIREPPNHTVTTEYWSQYATTTVTAPPGTDSVIREPPN 723
Db 1670 TTTTTPPTTTPSPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1725
QY 724 HVTVT-TEYWSQYATTTVTAPPGTDSVIREPPNPTVTTEYWSQYATTTVTAP- 781
Db 1726 TMTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1785
QY 782 -----PGTDTVII-----YESMSSKI-STSSNDI 806
Db 1786 CMTGWLDGCKPNFKPKG-DTELIGDVGCPGWAANTSCRATWPDVIGQGVTVCDV 1844
QY 807 T-----SIIP-SFSRPHVNTTSDISTFESSMNTPTSSDGMLLSTTL 852
Db 1845 SVGLICKEDQKPGVPMFCLNTEINQOCCEVT-OPITMTTNTENPTTPTTPTT 1903
QY 853 VTSEETTELICDKGKESRLSSSGVIVNPDNNESSIVTSTVPTASTMSDLSLSDGIS 912
Db 1904 VTPPTPTSTQSPGLQA-----PPTPTTSTTTTTPPTTPTGQTPTTPTT 1951
QY 913 ATSDNWSKGVSVTTTSV-----TTIQTTPNPLSSVTSLSQLSSIPSVSESKVTF 967
Db 1952 TTTTTPPTTPTGQTPTTTLITTTTMTPTTPTSTKSTVTPPTTPTTPTTPTT 2007
QY 968 TSNQDQSGTHDSQSTSEIIVTSTKVLPPVSSNTDLTSEPTNREQPTTSLTSTN 1027
Db 2008 -----PTGT-----QPTMTPISTTTT---VPTPTPTTGTGPTTSTAPIAELTTSN 2054
QY 1028 SITEDIT-----TSOPTGDNG-----DNTSSTNP-VPTVATSTLASASEEDNKS 1070
Db 2055 PPTSTPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 2104
QY 1071 GSHESA---STSLKP-----SMGNSGLTSTTEIATTTT-----PTEAPSPVSSG 1114
Db 2105 GGHITLSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 2163
QY 1115 TDVTTEPT-----DTRQD-----TTLSTTKNSLVAATQATNENCGKSPST 1158
Db 2164 TGLRPTSSVLLCCVLDNYIYAPGEEVNGYTGDTCTFVNCUSLCTLEFYWNPCSTPSP 2223
```

QY 1159 DLTSSLTGTGTASTANSSELVTSQS 1183  
Db 2224 TPTPSKSTPTPSKPSSTPSKPTPTGT 2248

## RESULT 12

T39174  
hypothetical Serine/threonine repeat containing protein [imported] - fission yeast (S  
C;Species: Schizosaccharomyces pombe  
C;Date: 03-Dec-1999 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
Accession: T50375; T39172; T39173; T39174; T39366  
R;Connor, R.; Churcher, C.M.; Wood, V.; Barrrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, February 1998  
A;Reference number: Z21832  
A;Accession: T50375  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1283 <CON>  
A;Cross-references: EMBL:AL021815; PIDN:CA861533.1; GSPDB:GN00067; SPDB:SPBC8E4.70c  
A;Experimental source: strain 972h-; cosmid c8E4  
A;Accession: T39172  
A;Molecule type: DNA  
A;Residues: 785-1283 <CO2>  
A;Cross-references: EMBL:AL021815; PIDN:CAA17000.1  
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL  
A;Accession: T39173  
A;Molecule type: DNA  
A;Residues: 'ME', 179, 'PLV', 183, 'W', 'KL', 556-761, 'HRGSS' <CO3>  
A;Cross-references: EMBL:AL021815; PIDN:CAA17001.1  
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL  
A;Accession: T39174  
A;Molecule type: DNA  
A;Residues: 1-555, 'S' <CO4>  
A;Cross-references: EMBL:AL021815; PIDN:CAA17002.1  
A;Note: this is an interim translation for a sequence replaced by GenBank/EMBL  
R;Wood, V.; Rajandream, M.A.; Barrrell, B.G.; Oliver, K.; Harris, D.  
submitted to the EMBL Data Library, March 1999  
A;Reference number: Z21848  
A;Accession: T39366  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-896, 'G', 898-904, 'I', 906-1283 <WOO>  
A;Cross-references: EMBL:AL035675; PIDN:CA838695.1; GSPDB:GN00067; SPDB:SPBC1289.15  
A;Experimental source: strain 972h-; cosmid c1289  
C;Genetics:  
A;Gene: SPDB:SPBC1289.15; SPDB:SPAC8E4.07c  
A;Map position: 2  
C;Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 8.9%; Score 578; DB 2; Length 1283;  
Best Local Similarity 25.6%; Pred. No. 1.1e-17;  
Matches 344; Conservative 170; Mismatches 571; Indels 258; Gaps 58;

QY 13 STAGAKITGVDFSNFTWSNAANYAFKPGCYPTWNAVGLSDGTSANPGDFTFLNMP 72  
Db 63 SFSNASVYTTTLITITNSQCTGNYSTLTIS--SQPLIHTNTSISKPSQATPONTNTQVS 120  
QY 73 CVFYKTTTSQTS-----VDLTADGVK---YATCFYSGEEFTTFTLCTV--NDALKSS 121  
Db 121 LTNGTTNYSNTSNLSLPTDITNGTTELIPTTSYNNQSHTLIYSTYTLPLNSTIDL 180  
QY 122 IKAQGTVTLPAPFNVGSGTSDELSKCTAGTNTVTENDGDKDISDVEFEKS----- 176  
Db 181 ILPHSTISTLSVINDTSAS--LSKTTSPTAGTITIVSG--SVGYSTFPASGTTSG 236  
QY 177 ---TWDPASVLYASVMSLKNKVTTLFVAPOCENYCTSGTMG-----ESSSN 220  
Db 237 TVEVVEPTAGTITIVSGSVGYSTTFPA---NGTSGIVVEVPEPTAGTITIVSGSV 292  
QY 221 GDAVIDCSNIHGITKGLNDWNP-----VSSSEFSYTKTCTNSGI---QIKYQNP 269

Db 293 GVTSTFEPAN---GTTSGTVEVVEPTAGTITETIVSGSVGYTSTFPANGTTSQTVEVVEPT 349  
 QY 270 AGYRPFDAIVATDYNQYTLAYTNDYTCAGSRLOSKPPTLRMTGYNKSDAGSNQIVIVA 329  
 Db 350 AG-----IVTETIVSG-SVGYTSTFPASGT-----TSQTVGVV 381  
 QY 330 --TTRTVD-----SPTAVTLPPNPVSDVKTKTIELOPIPTITITTSYVGVVTSVLTKT 382  
 Db 382 EPTAGTITETIVSGSKAFSTFPANGT--SGTVEVVEPTAGTITKIVSGSKT--FTST 437  
 QY 383 APIGETATVIVDPVHYHTTWTSE-WTGTI-TTITTRINPDSIDTVVQVPLPNTVST 440  
 Db 438 FPANGTTSQTVGVVEPTAGTITETIVSGSVGYTSTFPANGTTSQTVEVVE-----PTAGT 492  
 QY 441 TEYWSQSFATTTVTAPPGCT--DTVIIRPPNHVVT-TTEYWSQSFATTTVTAPPGCT 497  
 Db 493 ITETIVSGSKTFTFPASGTSTGTVVEVEPTAGTITETIVSGSKAFSTFPANGTSTG 552  
 QY 498 DSVIIRPPNPVTTEYWSQSFATTTVTAPPGTDS--VIRPPNPVTTEYWSQSF 555  
 Db 553 VEVV--EPTAGTITET--IVSGSVGYTSTFPASGTSTGTVVEVEPTAGTITET---IVS 604  
 - 556 YATTTVTAPPGTDS--VIRPPNPHTVTTTEYWSQSFATTTVTAPPGTDTVIRREP 613  
 Db 605 GSVGYTSTFPASGTSTGTVVEVEPTAGTITET--IVSGSVGYTSTFPASGTSTGTVVEVEP 663  
 QY 614 PNHVVT--TTEYWSQSFATTTVTAPPGCTDTVIRPPNPVTTEYWSQSFATTTTITA 672  
 Db 664 TAGTITETIVSGSKAFSTFPANGTTSQTVEV--EPTAGTITET--IVSGSKTFTSTFPA 720  
 QY 673 PPGTDTVLIREPPNHVTTEYWSQSFATTTVTAPPGCTDTVLIREPPNHVTTEYWS 732  
 Db 721 NGTTSQTVGVVEPTAGTITET--IVSGSVGYTSTFPASGTSTGTVVEVEPTAGTITETVI- 778  
 QY 733 SQSVATTTVTAPPGCTDTVIRPPNPVTTEYWSQSFATTTVTAPPGTDTVIRREP 792  
 Db 779 SGVSFNSITAHDTSSGAVIWEPTAGTITET--IVSGSIPFTSTIPAQGTTSQTVGVVE 837  
 QY 793 SMS---SKLSTSSNDITSLIPSRPHYVNSTDLSTFPSSNMNPTSISSDGMLLSS 849  
 Db 838 PTAGTITETIVSGSVGYTSTFPA-----QGTSGTVEVVEPTAGTITETIISGVGYT 890  
 QY 850 TLTVESET--TELICSDGKECRLSSSGIVTNPDSNEISIVTSVPTASTMSDSLSS 907  
 Db 891 STFPAGTTSQTVGVV-----APTAGTITETIVSGSIPFTSTIPAQGTTSQTVGV 940  
 QY 908 TDGISATSSDNWSKSVVTE-----TSVTIQTTPNPLSSVTSLTQLSSIPSVSESE 962  
 Db 941 VEPTAGTITETIISGVGYTSTFPAGTTSQTVGVV--PTAGTITETIVSGSIP----- 993  
 QY 963 SKVTFTNGDNQSGTHDSQSTSEIEIVTTSKVLPLPVVSSNTDLSEPTNTREOPTTL 1022  
 Db 994 ---FTSTIPAQ-----TTSQTVGVVEPTAGTITETIISGVGYTS---TTPAQGTTS 1040  
 QY 1023 ST-----TSNSITDITTSQPTGNDGNTSSNPNVPTVATST-----LASASEDNKS 1070  
 Db 1041 GTVEVVEPTAGTITETIVSG-----SIPFTSTIPAQGTTSQTVGVVPTAGTITST 1092  
 QY 1071 GSHESASTSLKPSMGNSGLTSTSEIATTTSTPASPVS-----GTD 1116  
 Db 1093 GSGTSWFTTVPATGRSG-----SVIVVPT---APACSTPPEACPNLNFPHSKN 1142  
 QY 1117 VTTEPTDTRQOPTTLSTSKTNSLVAATQATNENGKSPSTDLSTLTGTSAITSANS 1176  
 Db 1143 VNFPADIR---VISVTPKGNMYDATVO-----FTTSSTMSKKSLS 1181  
 QY 1177 ELVTSQVGTGAVASANDQSHSTSVTNSIVSNTPOTTLQQVTSSTSTNTFTAS-- 1234  
 Db 1182 ELKILG-----LSQTYLLYSYNSKVDN---ISNPGSWT--STVTQGTSTGSIYICMPH 1229  
 QY 1235 ---TYD---GSGSIIOHSTWLY 1250  
 Db 1230 PQIQYDNCAGVYDMSQCNWTSY 1252

RESULT 13  
 S57180  
 Probable membrane protein YJRL51c - yeast (Saccharomyces cerevisiae)  
 N: Alternate names: protein J2223; serine/threonine-rich protein YJRL51c  
 C: Species: Saccharomyces cerevisiae  
 C: Date: 23-Aug-1995 #sequence\_revision 08-Sep-1995 #text\_change 05-Nov-1999  
 C: Accession: S57180  
 R: Scarce, T.  
 Submitted to the Protein Sequence Database, September 1995  
 A: Reference number: S57169  
 A: Accession: S57180  
 A: Molecule type: DNA  
 A: Residues: 1-1161 <SCA>  
 A: Cross-references: EMBL:249651; NID:gl015902; PID:gl015903; GSPDB:GN00010; MIPS:YJRL  
 C: Genbank:  
 A: Gene: MIPS:YJRL51c  
 A: Map position: 10R  
 C: Keywords: transmembrane protein

Query Match 8.9% Score 577; DB 2; Length 1161;  
 Best Local Similarity 26.3%; Pred. No. 1e-17;  
 Matches 293; Conservative 131; Mismatches 427; Indels 264; Gaps 46;  
 QY 276 IDAYISATDVANQYTLAY-----TNDY--TCAGSRLOSKPPTLRMTGYNKSDAGSNGI 325  
 Db 36 IEIAYVVSIDIRAHIFQYYSFRNRHKTETYPSEIAAAVFDYGDFTRLTGISGDE----- 89  
 QY 326 VIVATRTVTDSTAVTTLFPNPSVDKTKTIE-ILQPIPTITITTSYVGVVTSYLTAKTAP 384  
 Db 90 ----VTRMITGVPMYSTRL--KPAISSALSXKDGIVTAIPTSTSTT-----TTKSSSTSTTP 138  
 QY 385 IGETATVIVDPVHYHTTTSQVTCGTTITTRINPDSIDTVVQVPLPNTVSTTEYW 444  
 Db 139 -----TTTTTS-----TTSTSTTPTTSTTST-----TPTTSTT--- 167  
 QY 445 SOSEATTTVTAPPGCTDTVIRPPNPVTTEYWSQSFATTTVTAPPGTDSVLIRES 504  
 Db 168 STPTTSTSTPTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTPT 227  
 QY 505 PNPNTVTTEYWSQSFATTTVTAPPGTDSVLIRESVLIRESVLIRESVLIRESVLIRES 564  
 Db 228 STPTTSTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTTSTA 272  
 QY 565 PPGTDSVLIRESVLIRESVLIRESVLIRESVLIRESVLIRESVLIRESVLIRESVLIRES 624  
 Db 273 PTST 317  
 QY 625 SOSFATTTVTGPPSGTDTVIRPPNPVTTEYWSQSFATTTVTAPPGCTDTVLIRES 684  
 Db 318 STTTFASLTTPATSTADTHT-----TSSSVSTNAFTSATTTT-----SDTVSSS 364  
 QY 685 PNPNTVTTEYWSQSFATTTVTAPPGCTDTVLIRESVLIRESVLIRESVLIRESVLIRES 734  
 Db 365 SPQSQTSSAEPTTVSEVTS--VEPTRSQVTSAP--TTVSEFTSSVPEPTRSQVT 418  
 QY 735 SYATTTTV---TAPPGCTDTVIRPPNPVTTEYWSQ-----SFATTTVTAP 781  
 Db 419 SSAEPTTVSEFTSSVPEPTRSQVTSAP--TTVSEFTSSVPEPTRSQVTSAP--TTVSEF 477  
 QY 782 PGCTDTVLIRESVLIRESVLIRESVLIRESVLIRESVLIRESVLIRESVLIRESVLIRES 834  
 Db 478 TSSVEPTRSQVTSAP--TTVSEFTSSVPEPTRSQVTSAP--TTVSEFTSSVPEPTRSQV 537  
 QY 835 NPFTSISDGMLLSSTLTVE---SETTT-----ELICSDGKECRLSSS 876  
 Db 538 TTPPEVSSFGSTSEITSSAEPLSPKATTSIAESSISNOITISSELIVS-----SVITSS 592  
 QY 877 SGIVTNPDSNEISIVTSV-PTA---SPMSDLSLSDGISATS----- 915  
 Db 593 SEIPSSIEVLTSSGSSSVPEPTSLVGPSSDESISSTESLSATSTFTSAVSSSKAADFTT 652



Db 1405 STTACQSTATNAS--SETPCNETQ--TSDKSTMTF 1439

RESULT 15  
S25345  
probable membrane protein YCR089w - yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein YCR1102  
C:Species: Saccharomyces cerevisiae  
C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 21-Jul-2000  
C:Accession: S25345; S19504  
R:Wilson, C.; Grisanti, P.; Frontali, L.  
yeast 8, 569-575, 1992  
A:Title: The complete sequence of a 6146 bp fragment of Saccharomyces cerevisiae chromos  
A:Reference number: S25345; MUID:9237594  
A:Molecule type: DNA  
A:Accession: S25345  
A:Residues: 1-1609 <WIL>  
A:Cross-references: GB:X59720; EMBL:S43845; NID:g1907116; PID:CAA42254.1; PID:g1907227  
R:Frontali, L.; Grisanti, P.  
submitted to the Protein Sequence Database, March 1992  
Reference number: S19504  
Accession: S19504  
A:Molecule type: DNA  
A:Residues: 1-1609 <PRO>  
A:Cross-references: EMBL:X59720; NID:g1907116; PID:e264634; PID:g1907227; MIPS:YCR089w  
C:Genetics:  
A:Gene: SGD:FTG2  
A:Cross-references: SGD:S0000685; MIPS:YCR089w  
A:Map position: 3p  
C:Keywords: transmembrane protein  
F:4-20/Domain: transmembrane #status predicted <TM1>  
F:1592-1609/Domain: transmembrane #status predicted <TM2>

Query Match 8.48; Score 547.5; DB 2; Length 1609;  
Best Local Similarity 22.58; Pred. No. 38-16; Indels 411; Gaps 66;  
Matches 332; Conservative 232; Mismatches 503;  
QY 55 SLDTGSANPGDTT-LNMPGVKTYTTSQVLDTPADGVKATCFQYGEFTFSLTCT 113  
DB 165 SLTFTVNPQSMTSPNSE-----KSALESIDFTSEISGTSF-KLESFDTGTITS 220  
QY 114 VND-----LKSSIKAF-----GVTPLPIAFNPGTGSST 143  
DB 221 YSPSPSKNSQNTLLSPLSLSSSGDLISLTIOATNDOGTKTPTLVDATSLPPT 280  
QY 144 DLEDSKCTAGTNTVFN-----DGDKDISIDVEFEKSTVDPSPAYL-----YASRVMP 191  
DB 281 LRSSMPTSGDSISHNFTSPKTSNGYDV-----LTSNIDPSLFTTSEYSTQLS 335  
DB 192 SLNKVT-----TLFVAPQCENGYT-SGTMGFSSNGDVAIDCSN 229  
DB 336 SUNRASKSETVNFSTASIASTPFGTDSATSLIDPISVSGSTASSFVGISTAFSTQNSNY 395  
QY 230 IHIGITKG---LNDW---NYPVS-----SESFSYTKTCTSNIGIQI 263  
DB 396 VPSTASGSSQYQDWSLSLPLSQTWWVINTNTGCVSTSTTSPAYVSTATKTVGVIT 455  
QY 264 KY-----QNVAGYRPFIDAYISATDNOYTLAYTNDYTCAGSRLOSKPFTLRWTG 314  
DB 456 EYVWCLPTQKQAIGVSSISVSPQASSFGSSILSSNSLTAAS--NNVPSTASGS 513  
QY 315 YKNSDAGSGIVATV-----RVTDSTTAVTLPNPSVDKTKTI-----EILQIPT 364  
DB 514 SQYQDWSLSLPLSQTWWVINTNTGCVSTSTTSPAYVSTATKTVGVIT 572  
QY 365 TTIITSYGVVTSYLNKT-----APIGETATVVDVPYHTTTTSEWGTITTTT 416  
DB 573 TQTKSQAIGISSITISATQSKPSILTIGISLQLSDATFKGTETIN---THLMTESTS 629  
QY 417 RTNPT-----DSIDIVVOVPLNPVSTTEYWSQSFATTTTAPGCTDVIIRPP 470  
DB 630 ITEPYFGTSDSFYLCYSEVNLAS--SLSSYPNFSSEGSTATII----- 673

Search completed: October 3, 2002, 15:33:15  
Job time: 252 sec



---

**THIS PAGE BLANK (USPTO)**



FT	DOMAIN	558	563	POLY-THR.	
FT	DOMAIN	594	599	POLY-THR.	
FT	DOMAIN	630	635	POLY-THR.	
FT	DOMAIN	666	671	POLY-THR.	
FT	DOMAIN	702	707	POLY-THR.	
FT	DOMAIN	738	743	POLY-THR.	
FT	DOMAIN	774	779	POLY-THR.	
FT	DOMAIN	874	877	POLY-SER.	
FT	CARBOHYD	471	471	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	579	579	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	615	615	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	687	687	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	723	723	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	820	820	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	886	886	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	918	918	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	973	973	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1045	1045	N-LINKED (GLCNAC. .)	(POTENTIAL).
FT	CARBOHYD	1068	1068	N-LINKED (GLCNAC. .)	(POTENTIAL).
SQ	SEQUENCE	1260 AA;	132641 MW;	763D1063A2354C24	CRC64;

Query Match 99.5%; Score 6465; DB 1; Length 1260;  
Best Local Similarity 99.6%; Pred. No. 5.3e-272;  
Matches 1255; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY	1	MLQOFTLLFLYLISIAAKITIGVDFSNLSLTWSNAAYAFKPGYPTWNAVGLWMSLDGTS	60
DB	1	MLQOFTLLFLYLISIAAKITIGVDFSNLSLTWSNAAYAFKPGYPTWNAVGLWMSLDGTS	60
QY	61	ANPGDFTLLNMPVFKYTTTSQTSVDLTADGVKATCOFYSGEEFTTSTLTCVNDALKS	120
DB	61	ANPGDFTLLNMPVFKYTTTSQTSVDLTADGVKATCOFYSGEEFTTSTLTCVNDALKS	120
QY	121	SIKAFGTVTLPIAFNNGVGTSGDLEDSKCFATGNTVTFNDGDKDISIDVEFEKSTVDP	180
DB	121	SIKAFGTVTLPIAFNNGVGTSGDLEDSKCFATGNTVTFNDGDKDISIDVEFEKSTVDP	180
QY	181	SAYLVASRVMPSLNKVTLFLFVAPCENGYSCTGMFSSSGNDVADCSNTHIGITKGLND	240
DB	181	SAYLVASRVMPSLNKVTLFLFVAPCENGYSCTGMFSSSGNDVADCSNTHIGITKGLND	240
QY	241	WNPVPSSEFSYTKTCTNSGTOIKQNPAGYRPFIDAVISATDVNQYTLAYTNDYTCAG	300
DB	241	WNPVPSSEFSYTKTCTNSGTOIKQNPAGYRPFIDAVISATDVNQYTLAYTNDYTCAG	300
QY	301	SPLOSKPFTLRWTKYKNSDAGSNGIVAVTRVTDSTTAVTTLFPNPSVDKTKTIELIQ	360
QY	301	SRQSKPFTLRWTKYKNSDAGSNGIVAVTRVTDSTTAVTTLFPNPSVDKTKTIELIQ	360
QY	361	PIPTTTTTSYGVVTSYLTKTAPIGETATVIVDVPYHTTTTSTVSEWGTGTTTTTRNMP	420
DB	361	PIPTTTTTSYGVVTSYLTKTAPIGETATVIVDVPYHTTTTSTVSEWGTGTTTTTRNMP	420
QY	421	TDSIDTWWVQVPLNPNTVSTTEYWSQSFAATTTVTAPPGGTDTVIIRPPNHTVTTTTEYW	480
DB	421	TDSIDTWWVQVPSNPNTVSTTEYWSQSFAATTTVTAPPGGTDTVIIRPPNHTVTTTTEYW	480
QY	481	SQSFAATTTVTAPPGGTDVIRPPNPTVTTTEYWSQSFAATTTVTAPPGGDSVIRE	540
DB	481	SQSFAATTTVTAPPGGTDVIRPPNPTVTTTEYWSQSFAATTTVTAPPGGDSVIRE	540
QY	541	PNPNTVTTTEYWSQYATTTVTAPPGGTDVIRPPNHTVTTTEYWSQYATTTVTVA	600
DB	541	PNPNTVTTTEYWSQYATTTVTAPPGGTDVIRPPNHTVTTTEYWSQYATTTVTVA	600
QY	601	PPGGTDVIRPPNHTVTTTEYWSQSFAATTTVTGPPSGTDTVIIRPPNPTVTTTTEYW	660
DB	601	PPGGTDVIRPPNHTVTTTEYWSQSFAATTTVTGPPSGTDTVIIRPPNPTVTTTTEYW	660
QY	661	SQSFAATTTVTAPPGGTDVIRPPNHTVTTTEYWSQSFAATTTVTAPPGTDVTLIRE	720
DB	661	SQSFAATTTVTAPPGTDVIRPPNHTVTTTEYWSQSFAATTTVTAPPGTDVTLIRE	720

QY	721	PNHHTVTTTEYWSQYATTTVTAPPGCTDVIIRPPNPTVTTTEYWSQSFAATTTVA	780
DB	721	PNHHTVTTTEYWSQYATTTVTAPPGCTDVIIRPPNPTVTTTEYWSQSFAATTTVA	780
QY	781	PGCGTDVIRPPNHTVTTTEYWSQSFAATTTVTAPPGCTDVIIRPPNPTVTTT	840
DB	781	PGCGTDVIRPPNHTVTTTEYWSQSFAATTTVTAPPGCTDVIIRPPNPTVTTT	840
QY	841	SDGMLLSSTTLVTESETTELICSDGKCSRLSSSGIVTNPDSNHSIVTSTVPTAST	900
DB	841	SDGMLLSSTTLVTESETTELICSDGKCSRLSSSGIVTNPDSNHSIVTSTVPTAST	900
QY	901	MSDSLSSGDSGATSDNVSKSGVVTETSVTTIQTTPNPSSSVTSLTQLSSIPSVSE	960
DB	901	MSDSLSSGDSGATSDNVSKSGVVTETSVTTIQTTPNPSSSVTSLTQLSSIPSVSE	960
QY	961	SESKVTFTSNGDNGSGTHDSQSTSTEIEIVTTSSTKVLPPVSSNTDLTSEPTNTREQPT	1020
DB	961	SESKVTFTSNGDNGSGTHDSQSTSTEIEIVTTSSTKVLPPVSSNTDLTSEPTNTREQPT	1020
QY	1021	TLSTTSNSITDITTSQPTDNGDNTSSTNPVPTVATSTLASASEDNKSGSHESASTSL	1080
DB	1021	TLSTTSNSITDITTSQPTDNGDNTSSTNPVPTVATSTLASASEDNKSGSHESASTSL	1080
QY	1081	KPSMGESGLTSTTEIEATTTPTPEAPSPAVSSGTDVTTPTDREQPTTLSTTSKTNSE	1140
DB	1081	KPSMGESGLTSTTEIEATTTPTPEAPSPAVSSGTDVTTPTDREQPTTLSTTSKTNSE	1140
QY	1141	LVATTOATNNGKSPSTDLTSLTGTSTASSELVTSGLVSGVGGAVASANDQSHST	1200
DB	1141	LVATTOATNNGKSPSTDLTSLTGTSTASSELVTSGLVSGVGGAVASANDQSHST	1200
QY	1201	SVTNSNSIVSNTPTLTSQQVTSSTSPSTNTFIASDYDGSIIQHSITWLYGLITLTLFLI	1260
DB	1201	SVTNSNSIVSNTPTLTSQQVTSSTSPSTNTFIASDYDGSIIQHSITWLYGLITLTLFLI	1260

RESULT 2  
ALS3\_CANAL STANDARD; PRT; 1119 AA.  
ID ALS3\_CANAL  
AC 074623;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Agglutinin-like protein 3 precursor.  
GN ALS3;  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RX MEDLINE=98309840; PubMed=9644209;  
RA Hoyer U.L., Payne T.L., Bell M., Myers A.M., Scherer S.;  
RT "Candida albicans ALS3 and insights into the nature of the ALS gene family.";  
RL Curr. Genet. 33:451-459(1998).  
CC -I- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
CC -I- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U87956; AAC39486.1; -;  
KW Cell adhesion; Glycoprotein; Repeat; Signal.  
FT SIGNAL 1 17



FT	CHAIN	18	1119	AGGLUTININ-LIKE PROTEIN 3.
FT	DOMAIN	433	792	10 X 36 AA TANDEM REPEATS.
FT	REPEAT	433	468	1-1.
FT	REPEAT	469	504	1-2.
FT	REPEAT	505	540	1-3.
FT	REPEAT	541	576	1-4.
FT	REPEAT	577	612	1-5.
FT	REPEAT	613	648	1-6.
FT	REPEAT	649	684	1-7.
FT	REPEAT	685	720	1-8.
FT	REPEAT	721	756	1-9.
FT	REPEAT	757	792	1-10.
FT	DOMAIN	399	404	POLY-THR.
FT	DOMAIN	430	435	POLY-THR.
FT	DOMAIN	557	563	POLY-THR.
FT	DOMAIN	593	597	POLY-THR.
FT	DOMAIN	630	635	POLY-THR.
FT	DOMAIN	666	671	POLY-THR.
FT	DOMAIN	702	707	POLY-THR.
FT	DOMAIN	738	743	POLY-THR.
FT	DOMAIN	774	777	POLY-THR.
FT	DOMAIN	1044	1047	POLY-THR.
FT	CARBOHYD	471	471	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	543	543	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	579	579	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	651	651	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	687	687	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	723	723	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	759	759	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	845	845	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	987	987	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1050	1050	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1061	1061	N-LINKED (GLCNAC. .) (POTENTIAL).
SO	SEQUENCE	1119 AA; 119927 MW; 6A3PB3FC8C879A71 CRC64;		
Query Match 64.7%; Score 4204.5; DB 1; Length 1119;				
Best Local Similarity 64.9%; Pred. No. 1.4e-174;				
Matches 844; Conservative 106; Mismatches 128; Indels 223; Gaps 18;				
Qy	1	MLQOFTLLFLYFASAKTITGVDSNSLTWSNANVAFKGCYPTWNAVGLNSLDGTS	60	
Db	1	MLQOYTLIIIVLSVAKTITGVDSNSLTWSNAAATYNYKPGCTPWNAVGLNSLDGTS	60	
Qy	61	ANPGDFTLNPCKVXYKTSQSDLTADGVKATCFQYSCSEBETTFSLCTVNDALKS	120	
Db	61	ASPGDFTLNPCKVXYKTSQSDLTADGVKATCFQAGEEFMTFLCTVNSLTLP	120	
Qy	121	SIKAGTFTLPIAPNCGTGSSTLSDSKCFTAGTNTVTFNDGDKOISIDVFEKSTVDP	180	
Db	121	SIKALGFTLPIAPNCGTGSSTLSDSKCFTAGTNTVTFNDGKKISINVDFFKSNVDP	180	
Qy	181	SAYLVASRWPSLNKVTILFVAPQENGNYTSGTGMGSSNGDVAIDCSNIHGIKGLND	240	
Db	181	KGLYDTSRVIPSLNKSVTILFVAPQACANGYSGTGMGFANTYGDVQIDCSNIHGIKGLND	240	
Qy	241	WNPVSSSEFSYTKTCTNSGIIQYQVNPAGYRPFDAIYISATDVNQYTLAYTNDYTCAG	300	
Db	241	WNPVSSSEFSYTKTCTSSNGIFITKVPAGYRPFVDAYISATDVNSYTLVANEYTCAG	300	
Qy	301	SRLOSKPFTLWTKYNSDAGSNGIVATRTVTDSTAVTTLPFNPVSDKTKTIELQ	360	
Db	301	GYWQAPFTLWTKYNSDAGSNGIVATRTVTDSTAVTTLPFDPNRDKTKTIELK	360	
Qy	361	PIPTTITTSVGVVTSYLTAKPIGETATVIVDPYHITTTVTSEMTGTTITTTTRNP	420	
Db	361	PIPTTITTSVGVVTSYLTAKPIGETATVIVDPYHITTTVTTSKWTCTIINSTHTNP	420	
Qy	421	TDSIDTVVQVPLNPNTVSTTEWWSQSFATTTVTPPGCTDTVLIREPNNHTVTTTEYW	480	
Db	421	TDSIDTVVQVPSNPNTVTTTEWWSQSFATTTIICPGCNTDTVLIREPNNHTVTTTEYW	480	
Qy	481	SQSFAATTTITAPPGCTDSDVLIREPNNPTVTTTEWWSQSFATTTITVTPPGCTDSDVIRE	540	
Db	481	SSSYTTTITFTAPPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITFTAPPGCTDSDVIRE	540	
Qy	541	PNPNTVTTTEWWSQSFATTTITAPPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVTA	600	
Db	541	PNPNTVTTTEWWSQSFATTTITAPPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVIA	600	
Qy	601	PPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVTPPGCTDSDVLIKEPPNPVTTTEYW	660	
Db	601	PPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVTPPGCTDSDVLIKEPPNPVTTTEYW	660	
Qy	661	SQSFAATTTITAPPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVTPPGCTDSDVIRE	720	
Db	661	SQSFAATTTITAPPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVTPPGCTDSDVIRE	720	
Qy	685	PNPNTVTTTEWWSQSFATTTITAPPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVTA	744	
Db	721	PNPNTVTTTEWWSQSFATTTITAPPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVIA	780	
Qy	745	PPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVTPPGCTDSDVLIKEPPNPVTTSSN	804	
Db	781	PPGCTDSDVLIKEPPNPVTTTEWWSQSFATTTITVTPPGCTDSDVLIKEPPNPVTTSSN	836	
Qy	805	DITSIIPSRPHVNSTTSDLSSTFESSMNTPTSISSDGMLLSSTLVTSEETTELIC	864	
Db	837	-----SFSRPHVNTHT-----LWSTTVIETKTITETSC	866	
Qy	865	SDGRCRLSSSGIVTPDGNESIVTSTVPTASTNSDLSSTDGISATSDNVKSGV	924	
Db	867	EGDKGCSMVSVSTRIVTPNNIETPMVNTVSTTTSTTS-QSPSCI-----FSESGV	918	
Qy	925	SVTTTTS-VTTIOTTPNLSSTVSLTQLSSIPSVSESKVTFSTNGDQSGTHDSQST	983	
Db	919	SVTETSSVTTAQTN-----PSVPTSEVTVTKGNENGPYESPST	961	
Qy	984	SFEIETVSTSTVLPVVSNTDLPSEPTNTREPTLTSTNSITEDITTSQPTGDCNG	1043	
Db	962	N-----VKSMDENSEPT-----	974	
Qy	1044	DNTSTNPVTVATSTLASAEEDNKSGSHESATSLKPMGENSLTSTFEI-----ATT	1100	
Db	975	-----TSTAAS-----TSTDENETIAT	992	
Qy	1101	TSPTFAPSPAVSGDVTTEPTDREOPTTLSTTKNSSELVATTOATNEN-GKSPSTD	1159	
Db	993	TGSVACGSPRISSADET-----TIVITTAESTSVI-----EPTNNKGGRAPSA-	1038	
Qy	1160	LTSLLTGTSASTANSELVTSQSVTGAVASANDQSHSTSVTNSNSIVSNVTPQTLTQ	1219	
Db	1039	-----TSPSTTTANNDSVITG-----TTNSQSQSQSYN-----SDTQQTLSQ	1080	
Qy	1220	QVTSSTSPSTNTFTASTYDGSIIQHSHTWLYGLITLTLSLFI	1260	
Db	1081	QMTSLVSLH--MLATFDGSGSVIQTSHWLCGLTLTLSLFI	1119	
RESULT 3				
ALAL	CANAL	ALAL	CANAL	STANDARD: PRT: 1419 AA.
DT	15-JUL-1999	(Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
DE	15-JUL-1999	(Rel. 38, Last annotation update)		
OS	Agglutinin-like protein ALAL precursor (Agglutinin-like adhesin).			
OC	ALAL OR ALS5			
OC	Candida albicans (Yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.			
PN	NCBI_TaxID=5476;			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98053977; PubMed=9393828;			
RA	Gaur N.K., Klotz S.A.;			

"Expression, cloning, and characterization of a Candida albicans gene, ALAL, that confers adherence properties upon Saccharomyces cerevisiae for extracellular matrix proteins.";  
 RL Infect. Immun. 65:5289-5294(1997).  
 CC 1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
 CC 1- PFM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch)  
 CC -----  
 CC EMBL: AF025429; AAB8883.1;  
 CC Cell adhesion; glycoprotein; Repeat; Signal.  
 KW SIGNAL 17  
 FT CHAIN 18 1419 AGGLUTININ-LIKE PROTEIN ALAL.  
 FT DOMAIN 399 404 POLY-THR.  
 FT DOMAIN 408 418 POLY-THR.  
 FT DOMAIN 437 441 POLY-SER.  
 FT DOMAIN 687 690 POLY-SER.  
 FT DOMAIN 700 703 POLY-SER.  
 FT DOMAIN 719 724 POLY-SER.  
 FT DOMAIN 749 752 POLY-SER.  
 FT DOMAIN 787 791 POLY-SER.  
 FT DOMAIN 869 872 POLY-SER.  
 FT DOMAIN 875 883 POLY-SER.  
 FT DOMAIN 901 911 POLY-SER.  
 FT DOMAIN 1216 1221 POLY-SER.  
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1301 1301 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1326 1326 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1419 AA; 149635 MW; 249F33F688A9D5B6 CRC64;

Query Match 49.2%; Score 3194; DB 1; Length 1419;  
 Best Local Similarity 49.5%; Pred. No. 6.6e-131;  
 Matches 708; Conservative 182; Mismatches 355; Indels 186; Gaps 29;

QY 1 MIQOFTLLFLYLSTAKITGVDFNSLITWSNAANYAFKPGYPTWNAVLCWSDGTS 60  
 DB 1 MIQOFTLLFLYLSTAKITGVDFNSLITWSNAANYAFKPGYPTWNAVLCWSDGTS 60  
 QY 61 ANPGDTFLNNPCVFKYTTTSQFSDTADGVKYATCOFYSGEFTTSLCTVNDALKS 120  
 DB 61 ANPGDTFLNNPCVFKYTTTSQFSDTADGVKYATCOFYSGEFTTSLCTVNDALKS 120  
 QY 121 SIKAGVTPLPIAFNNGCTGSDTLEDSKCTAGTWTVFNDGDKDISIDVEKSTVDP 180  
 DB 121 SIKALGVTLPIAFNNGCTGSDTLEDSKCTAGTWTVFNDGDKDISIDVEKSTVDP 180  
 QY 181 SAYLYASRVPLSKVTLFVAPCENGYSCTMGFSSNGDVAIDCSNTHIGITKGLND 240  
 DB 181 SGYLTSKFWPLSKVTLFVAPCENGYSCTMGFSSNGDVAIDCSNTHIGITKGLND 240  
 QY 241 WNPVSSSEFSYTKTCTNSGIIQYQNPAGYRPFIDAYISATVDNQYTLAYTNDYTCAG 300  
 DB 241 WNPVTSSEFSYTKSCSFGISITQNPAGYRPFIDAYISPDNNQYQLSKNDYTCVD 300  
 QY 301 SRLQSKPTFLRWTKYKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVDKTKTIEILQ 360  
 DB 301 DYWHQAPFTLKWTKYKNSDAGSNGIVIVATRTVTDSTTAVTTLFPNPSVDKTKTIEILQ 360  
 QY 361 PIPTTTITTSVGVGTTSYLTKTAPIGETATVVDVPYHTTTVTVTSEMTGTTTTTTRNP 420  
 DB 361 PIPTTTITTSVGVGTTSYLTKTAPIGETATVVDVPYHTTTVTVTSEMTGTTTTTTRNP 420  
 QY 421 TDSIDTVVQVPSNPNTTTTQFWESEFTSTTTITNSLKCTDSVIVREPHNPTVTTEFW 480  
 DB 421 TDSIDTVVQVPSNPNTTTTQFWESEFTSTTTITNSLKCTDSVIVREPHNPTVTTEFW 480

DB 421 TDSIDTVVQVPSNPNTTTTQFWESEFTSTTTITNSLKCTDSVIVREPHNPTVTTEFW 480  
 QY 481 SOSFATTTVTAPPGGDSVIVREPHNPTVTTEWWSQSFATTTVTAPPGGDSVIVRE 540  
 DB 481 SESFATTTITKPEGDSVIVREPHNPTVTTEWWSQSFATTTVTAPPGGDSVIVRE 540  
 QY 541 PNPPTVTTEWWSQSFATTTVTAPPGGDSVIVREPHNPTVTTEWWSQSFATTTVT 600  
 DB 541 PNPPTVTTEWWSQSFATTTVTAPPGGDSVIVREPHNPTVTTEWWSQSFATTT 600  
 QY 601 PGGDTVTIIRREPHNPTVTTEWWSQSFATTTVTAPPGGDSVIVREPHNPTVT 657  
 DB 601 GPEGDSVIVREPHNPTVTTEWWSQSFATTTVTAPPGGDSVIVREPHNPTVT 660  
 QY 658 -----EWSQSFATTTTAPPGGDSVIVREPHNPTVT----- 691  
 DB 661 SDSNLSAQSSESSVEQSSIVGLSSSDIPLSSDMPSSSGLTSESTVSSYSDSD 720  
 QY 692 -----TTEWWSQSFATTTVTAPPGGDSVIVREPHNPTVT----- 725  
 DB 721 SSSSTLSSDRCSISDITTFWSSSSDLESTITWSSSIDAQSASHLVOSVNSI 780  
 QY 726 VITTEWWSQSFATTTVTAPPGGDSVIVREPHNPTVTTEWWSQSFATTTVT 768  
 DB 781 STSQELSSSSSEESTF-----ATDALVSSOASSILSSDTSYSSSTISSDDFPH 835  
 QY 769 -----SOSFATTT-----TVTAPPGGDTVTIIRREPHNPTVT----- 806  
 DB 836 GEDSDLSISFITVEISSVSLATSDPASFDSSSLNDSSSSPSSDQSDILTSSFS 895  
 QY 807 TSIIPFS-----RPHVYNTTSDLTFFSSSMNTPTSISSDCKLLSSTLVTES 856  
 DB 896 TLVPSFSLSSSSLSLTYPHYVNTTTHASESSSVASPMASESAN--DDTYTLES 953  
 QY 857 ETTTELCSGDKG---CSRLLSSSSGIVTN-PDSNESSIVTSTVPTASTMSDLSLSDG-I 911  
 DB 954 TDTTSSIGDSTVTFCRRDNGDGCIVTGMPPSSSIDSEQTSDDVTTSFVASSTPTSAEQ 1013  
 QY 912 SATSDSNV--SKQGVSVTSTSTVTIQTTPNPLSSSVTSLTQLSSIPSVSESESKVFTS 969  
 DB 1014 SITDNPNDSSQTSKSSSVSDTVWNSILLSETS--TLSSDDSTS--SDTSISST 1070  
 QY 970 NGD---NOSGTHDSQSTSTETIEVITS---STKVLPPVYSSNTDLTSEPTNTR----- 1016  
 DB 1071 NEDTGNINAGSSHTSTASIKESIOKTGVLSSSYLSTKLSSTSDITIELITTEL 1130  
 QY 1017 -----EQTTLTSTNSITE-----DITTSQPTGDNDNTSTNPVPTVAT---STLA 1061  
 DB 1131 TTIEDNEPTTSTPSSSHSEISSONSVLKQVDRESTIKTPTTDDVTVSSLVHSTEA 1190  
 QY 1062 SA-----SEEDNKSCHESASTLAPSMGENSGLTSTETETATTSPTAPSPAVSSGTD 1116  
 DB 1191 STATLGNSFNVAFTSNATSLASTSSSNHATESGTVKSEASAIPTPTS--TD 1248  
 QY 1117 VTEPTDTRQPTTLTSTKTN-----SELVATTOATN---ENGCKSPSTDLTSSLTGT 1168  
 DB 1249 NRLSYSTEAKGITYANGSTNNLITESTQVAAPTSDSTVLIENPVVVTFTFDNSSAAVDQ 1308  
 QY 1169 SASTSANSLEY-----TSGSVTGCVASASNDOSHSTSV-----TNSNSI 1208  
 DB 1309 PKTKSIEESTMNPDTNETNNGFIATLSQAVPSSSIHSELISTTTAKTTDASMGDSA 1368  
 QY 1209 VSNTPQTLSQQVTSSTPNTFIATYDGGSGCIHQHSTWLYGLITLLSLF 1259  
 DB 1369 ANSQPTTLIOQVATS--SYNOPLITTVAGSSSATKHPWSLLKFLISVALFF 1417  
 RESULT 4  
 ID ALS2\_CANAL STANDARD; PRT; 468 AA.  
 AC 074657;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Agglutinin-like protein 2 precursor (Fragment).  
 GN ALS2.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1161;  
 RX MEDLINE=98440424; PubMed=9765564;  
 RA Hoyer L.L., Payne T.L., Hecht J.E.;  
 RT "Identification of Candida albicans ALS2 and ALS4 and localization of  
 RT als proteins to the fungal cell surface."  
 RL J. Bacteriol. 180:5334-5343(1998).  
 CC -|- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
 CC -|- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF024580; AAC64235.1; Repeat; Signal.  
 DR Cell adhesion; Glycoprotein; POTENTIAL.  
 KW SIGNAL 1 17  
 FT CHAIN 18 >468 AGGLUTININ-LIKE PROTEIN 2.  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 315 315 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT NON\_TER 468 468  
 FT SEQUENCE 468 AA: 50203 MW: BFE773E169ED0FAE CRC64;  
 SQ SEQUENCE 468 AA: 50203 MW: BFE773E169ED0FAE CRC64;  
 -----  
 Query Match 30.78; Score 1992.5; DB 1; Length 468;  
 Best Local Similarity 80.28; Pred. No. 1.2e-79;  
 Matches 376; Conservative 33; Mismatches 59; Indels 1; Gaps 1;  
 QY 1 MLOQFTLLLYLSIASAKTITGVDFSNLSWNAANYAFKPGYPTWNAVHGLSDGTS 60  
 DB 1 MLOQFTLLLYLSIASAKTITGVDFSNLSWNAANYAFKPGYPTWNAVHGLSDGTS 60  
 QY 61 ANPGDFTLLNMPGVKFTTQTSVDLTADGVKYATCFYSGEFTFTSLCTVNDALKS 120  
 DB 61 ANPGDFTLLNMPGVKFTTQTSVDLTADGVKYATCFYSGEFTFTSLCTVNDALKS 120  
 QY 121 SIKAFGTVTLPIAFNVGSGSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180  
 DB 121 SIKAFGTVTLPIAFNVGSGSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180  
 QY 181 SAYLYASRVNPLSNKVTTLFVAPQCENGYTSCTMGFSNGSDVAIDCSNIHIGITKGLND 240  
 DB 181 SAYLYASRVNPLSNKVTTLFVAPQCENGYTSCTMGFSNGSDVAIDCSNIHIGITKGLND 240  
 QY 241 WNPVSSSESFTYKTKCTSGNGIQKQNVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300  
 DB 241 WNPVSSDSLSYKTKCTSGNGIQKQNVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300  
 QY 301 SRLQSKPFTLRWTVGKNSDAGSNGIVVATRTVDTSTTAVTLPNPSVDKTKTEILQ 360  
 DB 301 SRLQSKPFTLRWTVGKNSDAGSNGIVVATRTVDTSTTAVTLPNPSVDKTKTEILQ 360  
 QY 361 PIPTTTITTSYGVGVTSYTKTAPIGETATVVDVPHHTTTTSEWGTGTTTTTTRNP 420  
 DB 361 PIPTTTITTSYGVGVTSYTKTAPIGETATVVDVPHHTTTTSEWGTGTTTTTTRNP 420  
 QY 421 TDSIDTVMVQVPLNPVTSTTEYWSQSFAITTTTAPPGGDTVIREP 469  
 DB 421 TDSIDTVMVQVPLNPVTSTTEYWSQSFAITTTTAPPGGDTVIREP 469

RESULT 5  
 ALS4 CANAL  
 ID ALS4 CANAL STANDARD; PRT; 469 AA.  
 AC 074650;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Agglutinin-like protein 4 precursor (Fragment).  
 GN ALS4.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1161;  
 RX MEDLINE=98440424; PubMed=9765564;  
 RA Hoyer L.L., Payne T.L., Hecht J.E.;  
 RT "Identification of Candida albicans ALS2 and ALS4 and localization of  
 RT als proteins to the fungal cell surface."  
 RL J. Bacteriol. 180:5334-5343(1998).  
 CC -|- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
 CC -|- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL; AF024584; AAC64239.1; Repeat; Signal.  
 DR Cell adhesion; Glycoprotein; POTENTIAL.  
 KW SIGNAL 1 17 AGGLUTININ-LIKE PROTEIN 4.  
 FT CHAIN 18 >469  
 FT NON\_TER 469 469  
 FT SEQUENCE 469 AA: 49604 MW: OEDCAB19B89EFCB1 CRC64;  
 SQ SEQUENCE 469 AA: 49604 MW: OEDCAB19B89EFCB1 CRC64;  
 -----  
 Query Match 28.08; Score 1819; DB 1; Length 469;  
 Best Local Similarity 72.18; Pred. No. 3.7e-72;  
 Matches 338; Conservative 48; Mismatches 83; Indels 0; Gaps 0;  
 QY 1 MLOQFTLLLYLSIASAKTITGVDFSNLSWNAANYAFKPGYPTWNAVHGLSDGTS 60  
 DB 1 MLOQFTLLLYLSIASAKTITGVDFSNLSWNAANYAFKPGYPTWNAVHGLSDGTS 60  
 QY 61 ANPGDFTLLNMPGVKFTTQTSVDLTADGVKYATCFYSGEFTFTSLCTVNDALKS 120  
 DB 61 ANPGDFTLLNMPGVKFTTQTSVDLTADGVKYATCFYSGEFTFTSLCTVNDALKS 120  
 QY 121 SIKAFGTVTLPIAFNVGSGSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180  
 DB 121 SIKAFGTVTLPIAFNVGSGSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDP 180  
 QY 181 SAYLYASRVNPLSNKVTTLFVAPQCENGYTSCTMGFSNGSDVAIDCSNIHIGITKGLND 240  
 DB 181 SAYLYASRVNPLSNKVTTLFVAPQCENGYTSCTMGFSNGSDVAIDCSNIHIGITKGLND 240  
 QY 241 WNPVSSSESFTYKTKCTSGNGIQKQNVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300  
 DB 241 WNPVSSDSLSYKTKCTSGNGIQKQNVAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300  
 QY 301 SRLQSKPFTLRWTVGKNSDAGSNGIVVATRTVDTSTTAVTLPNPSVDKTKTEILQ 360  
 DB 301 SRLQSKPFTLRWTVGKNSDAGSNGIVVATRTVDTSTTAVTLPNPSVDKTKTEILQ 360  
 QY 361 PIPTTTITTSYGVGVTSYTKTAPIGETATVVDVPHHTTTTSEWGTGTTTTTTRNP 420  
 DB 361 PIPTTTITTSYGVGVTSYTKTAPIGETATVVDVPHHTTTTSEWGTGTTTTTTRNP 420  
 QY 421 TDSIDTVMVQVPLNPVTSTTEYWSQSFAITTTTAPPGGDTVIREP 469

Db 421 TSDIDTVVQVSPNPTVTTEYNQSYATTTTATPPGGTDSVIIREP 469

RESULT 6  
FLOL YEAST STANDARD; PRT; 1537 AA.  
AC P32768;  
DT 01-OCT-1993 (Rel. 27, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Flocculation protein FLO1 precursor (Flocculin 1).  
GN FLO1 OR VAR050N.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
[1]  
SEQUENCE FROM N.A.  
BX MEDLINE=94262325; PubMed=8203162;  
Watarai J., Takata Y., Ogawa M., Sahara H., Koshino S., Onnala M.-L.,  
Aitkens J., Jaatinen R., Penttilae M., Keranen S.;  
"Molecular cloning and analysis of the yeast flocculation gene FLO1.";  
Yeast 10:211-225(1994).  
[2]  
SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RX MEDLINE=95249563; PubMed=7731988;  
Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,  
Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,  
Storms R.K.;  
"The nucleotide sequence of chromosome I from Saccharomyces  
cerevisiae.";  
Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).  
[3]  
PRELIMINARY SEQUENCE FROM N.A.  
RX MEDLINE=93289821; PubMed=8511970;  
Teunissen A.W.R.H., Holub E., van der Hucht J., van den Berg J.A.,  
Steenma H.Y.;  
"Sequence of the open reading frame of the FLO1 gene from  
Saccharomyces cerevisiae.";  
Yeast 9:423-427(1993).  
[4]  
REVIEW.  
RX MEDLINE=96076625; PubMed=7502576;  
Teunissen A.W., Steensma H.Y.;  
"Review: the dominant flocculation genes of Saccharomyces cerevisiae  
constitute a new subtelomeric gene family.";  
Yeast 11:1001-1013(1995).  
-!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.  
-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor  
(Potential).  
-!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).  
-!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL; X78160; CAAS5024.1;  
DR EMBL; J229520; AAC09499.1; ALT\_SEQ.  
DR FBL; S31230; S31230.  
DR SGD; S0000084; FLO1.  
DR InterPro; IPR001389; Flocculin.  
DR Pfam; PF00624; Flocculin; 18.  
DR GlycoProtein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 7 FLOCCULATION PROTEIN FLO1.  
FT PROPEP ? 1537 REMOVED IN MATURE FORM (POTENTIAL).

FT DOMAIN 274 1240  
FT FT 278 1087  
FT REPEAT 322 367  
FT REPEAT 323 367  
FT REPEAT 368 412  
FT REPEAT 413 457  
FT REPEAT 458 502  
FT REPEAT 503 547  
FT REPEAT 548 592  
FT REPEAT 593 637  
FT REPEAT 638 682  
FT REPEAT 683 727  
FT REPEAT 728 772  
FT REPEAT 773 817  
FT REPEAT 818 862  
FT REPEAT 863 907  
FT REPEAT 908 952  
FT REPEAT 953 997  
FT REPEAT 998 1042  
FT REPEAT 1043 1087  
FT DOMAIN 1118 1157  
FT REPEAT 1118 1137  
FT REPEAT 1138 1157  
FT DOMAIN 1226 1392  
FT REPEAT 1226 1276  
FT REPEAT 1291 1341  
FT REPEAT 1342 1392  
FT DOMAIN 1408 1434  
FT REPEAT 1408 1416  
FT REPEAT 1417 1425  
FT REPEAT 1426 1434  
FT CARBOHYD 135 135  
FT CARBOHYD 187 187  
FT CARBOHYD 262 262  
FT CARBOHYD 329 329  
FT CARBOHYD 374 374  
FT CARBOHYD 419 419  
FT CARBOHYD 509 509  
FT CARBOHYD 554 554  
FT CARBOHYD 599 599  
FT CARBOHYD 644 644  
FT CARBOHYD 689 689  
FT CARBOHYD 734 734  
FT CARBOHYD 1114 1114  
SQ SEQUENCE 1537 AA; 160722 MW; 992650C6BE9A8CEA CRC64;  
  
Query Match 14.5%; Score 940; DB 1; Length 1537;  
Best Local Similarity 26.0%; Pred. No. 1.1e-33;  
Matches 417; Conservative 244; Mismatches 523; Indels 418; Gaps 85;  
  
QY 2 LQQTLLFLYLSTASAKT-----ITGVDFDSF-----NSLTWSNAANYAP----- 40  
DB 10 LAVFTLLAL-TSVASGATEACLPAGORKSGMNFYQYSLKDSSTYSNAAYMAYGVASKT 68  
QY 41 -----KPGVPTM-----NAVLGWSLD--G 58  
DB 69 KLGVSVGOTDISIDYNI PCVSSSGTFPCPQEDSYGNMGCKMGACNSQGIAYWSTDLFG 128  
QY 59 TSANPCDPFTLNM-----PCVKYKTTSQ-----TSVD 85  
DB 129 FYTTFTNV-TLEMTGTVFLPQTGTSYTFKATVDDSAILLSVGGATAFNCCAQQPPTSIN 187  
QY 86 L7ADGVK-----YATCOFYSGEFTFTSTLCTVNDALKSSIKAFCT-----VTLP 131  
DB 188 FTIDGIRKPGWGSGLPPNIEGVTVMYAGYYPN-----KVYSNAVSGWGLPDISVTLP 238  
QY 132 IAFNVGGTGSSTDLDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTV-DPSAYLIASRYM 190  
DB 239 -----DGTVTSDDEG-----YVSEFD-----DLSOSNCTVPPDSNYA-VSTTT 277  
QY 191 PSLNKVTTFLVAFQCENGYTSMTGMFSSNSGNDVAIDCSNIHIGITKGLNDNPNVPSSESF 250

```
Db 278 TTTEPWTGTTSTSTEMTTVTGTCNGVPTDETVIVIRPTTASTIIITTEPMNSTFTSTST 337
Qy 251 SYTKTCSNGIQKYNQVAGYRPFIDAYISADVNQVNTLAYINDYTCAGSLQSKPFTL 310
Db 338 ELTTVTGTCNGVPTDETVII-----VIRPTTATTAITTEPMNSTFTSTSTEL----- 384
Qy 311 RMTGKNSDAGSNGI---VIVATRTVDSITTAATL--PENSVDKTKT----- 355
Db 385 -----TTVTGTCNGVPTDETVIVIRPTTATTAITTTQPNNDTFTSTEMTTVTGTNGL 338
Qy 356 -----IEIQPTPTTIT-----SYGVGTSYLVKTKA---PIGATAVIVDVPYH 438
Db 439 PTDETVIVIRPTTATTAITTTQPNNDTFTSTEMTTVTGTNGLPTDET--IIIVIRPTT 497
Qy 399 TTT--TTVSEMTGTTT-----TTTTTN--PTDSIDVAVVQVPLNPT-VSTTEYWSQS 447
Db 498 ATTAMTTQPNNDTFTSTEMTTVTGTCNGVPTDE--TIIVIRPTTATTAITTTQPNNDT 556
Qy 448 FATT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTEYWSQSFAIT-----TTVTAPP 494
Db 557 FSTSTEMTTVTGTCNGVPTDETVIIIVIRPTTATTAITTEPMNSTFTSTSTELTVITGN 616
Qy 495 G-GTDS--VIIRPPNPT--VTTEYWSQSFAIT-----TTVTAPPG-GTDS--VIIRPP 542
Db 617 GLPTDETVIVIRPTTATTAITTTQPNNDTFTSTEMTTVTGTNGLPTDETIIIVIRPT 676
Qy 543 NPT--VTTEYWSQSFAIT-----TTVTAPPG-GTDS--VIIRPPNHT--VTTEYWSQS 591
Db 677 TATTAMTTQPNNDTFTSTEMTTVTGTCNGVPTDETVIIIVIRPTTATTAITTTQPNNDT 736
Qy 592 YATT-----TTVTAPPG-GTD--TVIIRPPNHT--VTTEYWSQSFAIT-----TTVT 635
Db 737 FSTSTEMTTVTGTCNGVPTDETVIIIVIRPTTSEGLISTTTEPMGTFTSTSTEMTTVTG 796
Qy 636 GPPSGTDTVIIRPPNPT--TVTTEYWSQSFAIT-----TTITAP--PGETDTVLIREPP 686
Db 797 GQPTDETVIVIRPTSEGLITTTTEPMGTFTSTSTEMTTTGTGTCNGVPTDETVIVIRPT 856
Qy 687 NHTV--TTTEYWSQSFAIT-----TTVTAPGE--TD--TVIIRPPNHTV--TTTEYWSQS 735
Db 857 SEGLISTTTEPMGTFTSTSTEMTTTGTGTCNGVPTDETVIVIRPTSEGLISTTTEPMGT 916
Qy 736 YATT-----TTVTAPPG-GTD--TVIIRPPNPTV--TTTEYWSQSFAITTTTAPPGGT- 785
Db 917 FSTSTEMTTVTGTCNGVPTDETVIVIRPTSEGLISTTTEPMGTFTSTSTSTVITGTN 976
Qy 786 -----DTVIIRPPNPT-----FTSSNDITSIIPFSRP----- 816
Db 977 GQPTDETVIVIRPTSEGLISTTTEPMGTFTSTSTEMTTTGTGTCNGVPTDETVIVIRPT 1036
Qy 817 --HYVNSTSD-LSTFESSMNTPTSISSDGLLSSTTLVTESETT-----TELICSDGKE 869
Db 1037 SEGLITTTTEPMGTFTSTSTEMTTTGTGTCNGVPTDETVIVIRPTTAISSSSSSSGQI 1096
Qy 870 CSRLLSSSGIVTNP--DSNESSIVTNPASTMSDLSLSDGTSISADNVSQVSVT 927
Db 1097 TSSITSSRPIT--PFTPSNGTSSVSSSVSSVTSLSFTSPVSSSVSSSTTTSIF 1155
Qy 928 TETSVTITQTTNPLSSVTSLSLSPSVSESKVTFTSNGDNQSDHDSOSTSTEI 987
Db 1156 SESSKSSV--IPTSSSTSSGSESTSSAGSVSS-----SFTSSSESKSPYSSSS---L 1205
Qy 988 EIVTSSST-----KVLPPVVSNTDLTSEPTNTRQPTTLSTTS-----NSITEDIT 1035
Db 1206 PLVTSATTSQETASSLPAPT-----TKTSEQTTLVTVTSSCHVCTESIPAVS 1256
Qy 1036 SQPTGNDGNTSSTN--PVPVTATSTLASSEDNKGSHESASTSLKPSMEGSLGTS 1093
Db 1257 TATVTVSGVTEVTTWCPSTSTETTKTKTBTQETTKOTTVTIVISSCESDVCSTAS 1316
Qy 1094 TEATVATTSPEAPSAPVSSGVDVTEPTDTREOPTTLSTT-----SKTSELVATT- 1145
Db 1317 PAIVSTSTATINGVTTEYTTWCPIST--TESRQITTLVTVTSSCGVCSSTASPAIVSTA 1374
```

Qy 1146 -----QATNENGKSPSDDLTSLLTGT--SASTSANSSELVTSQVGTGA 1188

Db 1375 TATVNDVTVPTWROTANESVSSKMSATGETTTNTLAETTTNTVAETINTCAA 1434

Qy 1189 -----VASASNDOSHSTV-----TNSNIVSNTPTTLTSLQVTS 1224

Db 1435 ERKTVTSSILSRNHAETOTASATDVIGHSSVSVSETGNTKSLTSSGLSTMSQPRST 1494

Qy 1225 SP-----SNTFIATVDGSGSIIOHSTWLYGLITLLSLFI 1260

Db 1495 PASSVGVSTASLEISTAGSANSLLAGSLSVTFASLLAI 1536

RESULT 7

YAG3\_YEAST

ID YAG3\_YEAST STANDARD; PRT; 1322 AA.

AC P39712-1995 (Rel. 31, Created)

DI 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE Hypothetical 138.1 kDa protein in FLO9-GDH3 intergenic precursor.

GN YAL063C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RX MEDLINE=95249563; PubMed=7731988;

RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,

RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,

RA Storms R.K.;

RT "The nucleotide sequence of chromosome I from Saccharomyces

RT cerevisiae."

RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).

RC -I- SIMILARITY: STRONG, TO YEAST PROTEIN FLOI.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

CC -----

DR EMBL; U12980; AAC04971.1; -

DR SGD; S0000059; YAL063C

DR InterPro; IPR001389; Flocculin.

DR Pfam; PF00624; Flocculin; 13

KW Hypothetical protein; Glycoprotein; Transmembrane; Signal.

FT SIGNAL 1 24

FT CHAIN 25 1322

FT TRANSMEM 366 388

FT TRANSMEM 754 775

FT CARBOHYD 135 135

FT CARBOHYD 175 187

FT CARBOHYD 203 203

FT CARBOHYD 257 257

FT CARBOHYD 262 262

FT CARBOHYD 270 270

FT CARBOHYD 329 329

FT CARBOHYD 419 419

FT CARBOHYD 464 464

FT CARBOHYD 509 509

FT CARBOHYD 554 554

FT CARBOHYD 599 599

FT CARBOHYD 644 644

FT CARBOHYD 689 689

FT CARBOHYD 734 734

FT CARBOHYD 888 888

FT SEQUENCE 1322 AA; 138072 MW; ADPDLF13267CEA CRC64;



CC -1- SUBUNIT: MULTIMERIC.  
CC -1- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,  
CC BRONCHUS, CERVIX AND GALL BLADDER.  
CC -1- PTR: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND  
CC VARIES AMONG DIFFERENT ALLELES.  
CC -1- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT  
CC OF SILKWORM HEMOCYTIN.  
CC -1- SIMILARITY: CONTAINS 2 VNFC DOMAINS.  
CC -1- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; L21998; AAB95295.1; -;  
DR EMBL; M74027; AAS59875.1; -;  
DR EMBL; M94131; AAS59163.1; -;  
DR EMBL; M94132; AAS59164.1; -;  
DR WIM; 158370; -;  
DR InterPro; IPR000359; Cys\_knot.  
DR InterPro; IPR000561; EGF-like.  
DR InterPro; IPR002400; GF\_Cys\_knot.  
DR InterPro; IPR001007; VNFC.  
DR InterPro; IPR001846; Vwd.  
DR Pfam; PF00007; Cys\_knot; 1.  
DR Pfam; PF00094; vwd; 4.  
DR PRINTS; PR00438; GFCYSKNOT.  
DR SMART; SM00214; VNC; 2.  
DR SMART; SM00216; VWD; 4.  
DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
DR PROSITE; PS0185; CTCK\_1; 1.  
DR PROSITE; PS01225; CTCK\_2; 1.  
DR PROSITE; PS01208; VNFC; 2.  
DR Glycoprotein; Repeat; signal.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 1 5179 MUCIN 2.  
FT DOMAIN 1401 1747 APPROXIMATE REPEATS.  
FT REPEAT 1401 1416 1.  
FT REPEAT 1417 1432 2.  
FT REPEAT 1433 1448 3.  
FT REPEAT 1449 1464 4.  
FT REPEAT 1465 1471 5.  
FT REPEAT 1472 1478 6.  
FT REPEAT 1479 1494 7A.  
FT REPEAT 1495 1517 7B.  
FT REPEAT 1518 1533 8A.  
FT REPEAT 1534 1556 8B.  
FT REPEAT 1557 1572 9A.  
FT REPEAT 1573 1596 9B.  
FT REPEAT 1597 1612 10A.  
FT REPEAT 1613 1635 10B.  
FT REPEAT 1636 1651 11A.  
FT REPEAT 1652 1675 11B.  
FT REPEAT 1676 1693 12.  
FT REPEAT 1684 1699 13.  
FT REPEAT 1700 1715 14.  
FT REPEAT 1716 1731 15.  
FT REPEAT 1732 1747 16.  
FT DOMAIN 4815 4886 VNFC 1.  
FT DOMAIN 4924 4991 VNFC 2.  
FT DOMAIN 5075 5160 CTCK.  
FT DISULFID 5075 5122 BY SIMILARITY.  
FT DISULFID 5089 5136 BY SIMILARITY.  
FT DISULFID 5098 5152 BY SIMILARITY.  
FT DISULFID 5102 5154 BY SIMILARITY.  
FT DISULFID 5159 ? BY SIMILARITY.

FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 670 670 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 894 894 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1139 1139 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1154 1154 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1215 1215 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1230 1230 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1246 1246 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1787 1787 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1820 1820 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4339 4339 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4351 4351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4362 4362 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4373 4373 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4422 4422 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4438 4438 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4502 4502 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4616 4616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4627 4627 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4752 4752 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4787 4787 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4881 4881 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4888 4888 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4955 4955 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 4970 4970 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 5019 5019 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 5038 5038 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 5069 5069 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CONFLICT 1351 1351 H -> L (IN REF. 3).  
FT CONFLICT 1412 1412 T -> S (IN REF. 3).  
FT CONFLICT 1449 1449 L -> P (IN REF. 3).  
FT CONFLICT 1504 1504 M -> T (IN REF. 3).  
FT CONFLICT 4192 4192 G -> S (IN REF. 2).  
SQ SEQUENCE 5179 AA; 540295 MW; 85CD7571FB9A5663 CRC64;  
  
Query Match 11.1%; Score 721.5; DB 1; Length 5179;  
Best Local Similarity 25.3%; Pred. No. 1.le-23;  
Matches 335; Conservative 158; Mismatches 554; Indels 279; Gaps 54;  
  
Qy 59 TSANFGDTFTLNMPGVFKYTTTSQTSVDLTADGVKATCOFYSGEEFTTFTLTCTVNDAL 118  
Db 1647 TTTTTPPTTTPSP-----TTTTSPPTTTTTPPTT-----TPSSPTTTPSPPT 1693  
  
Qy 119 KSIKARGTVTLPPIAFNVGGTGSSTDLSDSKCTAGTNVTFTNDGDKDISIDVEFEKSTV 178  
Db 1694 TMTTTPPTTTPSSPTTTTTPSTTTPSPPTTMTTTPSTTTPSPPTTMT-----TTL 1747  
  
Qy 179 DP---SAYLYASRVMSLNKVT-----TLFVAPQCE-NGY-TSGTMGFSSSDGVA 224  
Db 1748 PTTTSSPLTTTTPSTTTPPTTTPPTTTPPCVPLCNWTKWLDGKPNHFKPGDTE 1807  
  
Qy 225 IDCSNIHIGITKLDWNPVSSSEFSY-----TKYC-TSGIOIKYQN-VPAGY 272  
Db 1808 L-----IGDVC8-PGWAANISCRATWYPDVIGOLGVVCDVSVGLICRNEQKEGV 1860  
  
Qy 273 REFIDAYISATDVNQYTLAYTNDYTCAGSLQSKPFTLRMTGYKNSDAGSNQIVATPR 332  
Db 1861 IFM-----AFCLNYEINQCC--ECVTQPTMTTNTTENPTPTTPTTPTTTT 1907  
  
Qy 333 TVDSTTAVTLPNPNVDKTKTIEILOPITTTTTSYGVVTSYLTAKTAPIG-ETATV 391  
Db 1908 TPTPTGTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1967  
  
Qy 392 IDVPYHTTTTTSWTGT-----ITTT-----TRTNPTDSIDTVVQVP 432  
Db 1968 ---PTITTTTTPPTTPTTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2024  
  
Qy 433 LPNPTVSTTEYWSQSFATTTTVPAGGTDTVIIEPPNHTVTTTEYWSQSFATTTVA 492  
Db 2025 TPTPTGTGTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2072





```
Db 184 TSTNFTI--NGIKPMD-GSLPDNITGVYMYAGYYPLKVVYNAVSMGTLPI--SVE-- 236
Qy 354 KTIIELOPIPT-TTITTSVGVVTSYLTAKTAPIGETATVIVDVPYHTT---TTVTSWMTG 409
Db 237 -----LPGDGVVSDNFEGVYVSE---DDDLQSQNCITPDPISIHTTSTITTTTETWIG 285
Qy 410 TIT-----TTTTRNPTDSIDTVVQVPLPNTVSTTXYWSQSFATTTVTPPGGTD 463
Db 286 TTTSTSTEMTITDNGQLTDEIV---IRPT-----TASTIT----- 322
Qy 464 VIIREPPNHVTTTXYWSQFAT---TTVTAPG-GTDS--VIIREPPN--TVTITE 514
Db 323 -----TTTETWIGTISTEMTIVTGTNGOPTDETIVIRTPSEGLITTTT 371
Qy 515 YWSQFAT---TTVTAPG-GTDS--VIIREPPN--TVTTEYWSQYATTTVTP 565
Db 372 PWGTFTSTEMTIVTGTNGOPTDETIVIRTPSEGLITTTTETWIGTFTSTEVTT 431
Qy 566 PGDTSVIIREPPNHVTTTXYWSQYATTTVTPPGGTDVIRPPNH--TVTTEY 623
Db 432 ITGTNG---OPTDETIV-----IVIRTPSEGLITTTTET 462
Qy 624 WQSQFAT---TTVT---GPPSGTDVIRPPNPV---TTTEYWSQYAT---TTI 670
Db 463 WGTFTSTSTEMTIVTGTNGOPTDETIVIRTPSEGLITTTTETWIGTFTSTEVTTI 522
Qy 671 TAPGE-TD--TVLIREPPNH--TVTTEYWSQYAT---TTVTAPGE-TD--TVLI 718
Db 523 TGTNGOPTDETIVIRTPSEGLITTTTETWIGTFTSTEMTIVTGTNGOPTDETIVI 582
Qy 719 REPPNH--TVTTEYWSQYATTTVTPPGT-----DTVIREPPNPVTVTTEYWSQ 770
Db 583 RPTSEGLITRTTETWIGTFTSTEVTTITGTNGOPTDETIVIRTP--TAISSLS 641
Qy 771 SFATTTTVA-----PPGGTDV---IYESMSKSISSNDITSIPFSRPHY 818
Db 642 SGQITSSITSRIIPFPVPSNGTSVSSSVSSVSSVSSVSSVSSVSSVSSVSSVSS 695
Qy 819 VNSTSDLSSTFESSNMT--PTSSSDGMLSLTLVTESETTELICSDGKESRLSS 876
Db 696 --STTTSIFSSTSSVPTSSSTSGSESKTSSASSSSSSSSSSSSSPKSPTNSSSS 753
Qy 877 SGIVTPDSESSIVTSTVPTATMSDLSLSDGLSAT---SDNVKSGVSVTTET 930
Db 754 LPVTSATTCQET--ASSLPATTTKTS-EQTLVTVTSCSHVCTESISSAIVSTATVT 810
Qy 931 --SVTTTOTTNPPLSSVTSLSQTSIPVSESESKVTFTSNGDNGSTHDSQSTSEIE 988
Db 811 VSGVTETWTCFISTET-----KQTKGTTGTQKGTETQ 847
Qy 989 IVTTSKVLPPVSVSNDLTSEPTNTRQOPTLTSTNSITEDITSTQDNGDNTSS 1048
Db 848 TETTKQTVV--TSSCEDICKTAS---PAIVSTSTATI-----NGVTTEY 890
Qy 1049 TNPVTATSLASEEDNKSGHESASTSLKPSMGENGSLTTEIETATTPTEAP- 1107
Db 891 TWCPISTTES-----KQTLVTVTSCSVCSEGVSETTSPA 925
Qy 1108 --SPAVSSGTDVTT-----EPTDTRQOPTLTSTSKTNSELVATQATNENGKSPSTD 1160
Db 926 IVSTATATVNDVVVTVPTWRPQTNEQ---SVSSKMNS---ATSETTNTGAETATV 978
Qy 1161 TSSLTGTSTASTSANSSELVTSQVTCGAVASANDQSHSTSV-----TNSNSIVSNTPQT 1215
Db 979 TSSLSRFNHAETQ-----TASATDVIHSSVSVSVSETGNTWMTSLSGLS 1023
Qy 1216 TLSQOVTS-----SPSTNPIASTYDGSIGIIQHSITWLYGLITLLSLFI 1260
Db 1024 TWSQOPRSTPASSMVGSSSTASLEISTAGSANSLLAGSLSVFIASLLAI 1074

RESULT 10
AMYH_YEAST
```

```

ID AMYH_YEAST STANDARD: PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucosylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-
GN glucosylase (1,4-alpha-D-glucan glucohydrolase)).
GN STAI OR STA2 OR MAL5 OR YR019C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
(1)
SEQUENCE FROM N.A.
RC STRAIN=S488C / AB972;
RA Barrrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
(2)
SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RP MEDLINE=87194600; PubMed=3106330;
RA Yamashita I., Nakamura M., Fukui S.;
RT "Gene fusion is a possible mechanism underlying the evolution of
RT STAI.";
RL J. Bacteriol. 169:2142-2149(1987).
(3)
SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes
RT from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
CC glucose residues successively from non-reducing ends of the chains
CC with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPCC215.13
CC -1- SIMILARITY: SOME TO S.POMBE SPCC285.13C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by, non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z38061; CAA86176.1; -;
CC EMBL; M16164; AAA35014.1; -;
CC EMBL; M16165; AAA35015.1; -;
CC EMBL; X13857; CAA32069.1; -;
CC PIR; B26877; B26877.
CC PIR; A26877; A26877.
CC PIR; S48478; S48478.
CC SGD; S0001458; MUC1.
CC Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
CC Signal; Multigene family.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 1367 GLUCOMYLASE S1/S2.
CC DOMAIN 210 1367 SER/THR-RICH.
CC FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

Query Match 10 48; Score 675.5; DB 1; Length 1367;
Best Local Similarity 26.28; Pred. No. 2.5e-22;
Matches 378; Conservative 177; Mismatches 535; Indels 353; Gaps 65;
```





QY 230 IHGIRKGLNDW---NYPVS-----SEPSYKTCISNGIOI 263  
Db 396 VPESTAGSQYODMSSSLPLSQTQWVINTNTOGSVSTTSPAVTATKVDGVIT 455  
QY 264 KY-----QNPAGVRPFIDAYISATDQVGYTLAYTNDYTCAGRLQSKPFLRWG 314  
Db 456 EVTVKPLQTKQAIGVSSSISSVQASFSGSSILSSNSTLAAS--NNVPSTFASG 513  
QY 315 YKNSDAGSNGIVAVT-----RVTDSATVATLFPNPSVDKTKTI-----EILQPIPT 364  
Db 514 SOYQDWSSSSLPLSQTQWVINTNTOGSVSTTSPAVYST-ATKTVGVITEYVWCP 572  
QY 365 TTITTSVGVVTSYLTKT-----APIGETATVIVDPVHTTTTSEMTITITTT 416  
Db 573 TOTKQAIGISSTISATQTKSPSILLTGISTLQSLDATFKGTETIN---THLMTST 629  
QY 417 RINPT-----DSIDRVVQVPLNPVSTTEYWSQSFATTTTVPAGTGDVIREPP 470  
Db 630 ITEPTFGTSDSYLCTSEVNLAS-SLSSYPNFSSEGSTATIT-----673  
QY 471 NHTVT---TTEYWSQSFATTTVTA-----PPGCTDSVIREPPNPTVTTREY 515  
Db 674 NSTVTFGSTSKYPSVSNPTEAQSQVSSVNSLTDFTSNSTETIAVISNIHKTSSNDY 733  
QY 516 WSQSFATTTTVPAGTGDVIREPPNPTVTT-----TEX--W--SQSYATTTVAP 565  
Db 734 -----SLTTQLKTSQKQILV-----STVTTVNGAATEITWCPASSIATVTSI- 781  
QY 566 PGGTDSVIREPPNHTVTTTEYWSQSFATTTTVPAGTGDVIREPPNHTV---TTE 622  
Db 782 -----KTLVLTEVCHSSECTPTVITSVATSTPIPLSTSSSTVLSSTVSE 828  
QY 623 YWSQSFATTTTVPAGTGDVIREPPNPTVTTTEYWSQSFATTTTITAPPGETDVL 682  
Db 829 CAKPAASEVINTQVSATS-----EATSTQ-----VSATATASESTTSQ 874  
QY 683 REPPNHTVTTTEYWSQSFATTTTVPAGTGDVIREPPNHTVTTTEYWSQSFAT 738  
Db 875 VSTASETIST--LGTONFTTGLSLFLPALSTEMINTVVSRTK---LIITEVCHS 929  
QY 739 TTVT-----APPGG-----TDTVIREPPNPTVTTTEYWSQSFATTTTVPAG 783  
Db 930 PTVITEVVTYKSGTPSNHSSQTLQTEAVEYTLSSHQVTMTSEVCSNISCTPVITSQ 988  
QY 784 GTDVTIYESMSSSKISSTSSDIISIIPSRPHYVNSTDLSTFESSMNPPTSISD 843  
Db 989 MRSTPPFPLTSTSSSLASTKSSL-----EASSEMSTFVSQSLAFTCS 1037  
QY 844 GMLLSSTLVTESETT--TELICSDGKCSRLSSSGIVTN--PDSNES-----SIV 891  
Db 1038 EK--RSTTSVSQMSNTVLNTIMS-----SSSNVISTNEKPSSTSPYNFSSGYS 1087  
QY 892 NSTVPTASTMSLSLSDGI-----SATSDNVS 920  
Db 1088 SSSPTSQYSLSTATTINGIKTVTWCPLEKSVTAASSQSSRSVDRFVSSSKPSSLS 1147  
QY 921 KSGVSVTTSVTIO-----TTPNPLSSVT--SLTQLSISVSESESKVFTSNGD 972  
Db 1148 QTSQYTLSTATTIISGLKTVTWCPLETSKSLGATQTS-----STAKVRIYS-AS 1199  
QY 973 NOSQTHSQSTSEIIVTSTKVLPPVSSWDLTSE-PTNREOPTLS-----TTS 1026  
Db 1200 SATSTLSLSTSESE-----SSSGYLSKGVCSGCTEQDQVQSSSPASTLAYSPSV 1255  
QY 1027 NSITEDITTSQPTGDNDGNTSTNPVTATSTLASASEDNKSGSHESASTSLKPSMGE 1086  
Db 1256 SSSSFTSTTASTL-----TSTHTSVPLPSSSSISASPSSTSS-----LLSTLS 1305  
QY 1087 NSGLTTEATEATTSPTEAPASVSGTDTVTEPTDREOPTLSLSTKTNSEL----- 1141  
Db 1306 SSTLPTATAVSSSTFTASSLP-----LSSKSLSLSPVSSSILMSFOSSSSSSSLAS 1362  
QY 1142 -----VATTQATNENGCKSPST-----DLTSSLT-----TGTS 1169

Db 1363 LSISPTVDTVSVLQPTTSLATLCTDSOCQOEVSITCNSCNDSTATTPSTVDTM 1422  
QY 1170 ATSANSELVTSVGTVG-----GAVASASND--QSHSTSVTNS-----NSIVS 1210  
Db 1423 TCTGSECQKTTSSSCDGYCKVSEYTKSSATISACSGEGCOASATSELNSQYVTMTSVIT 1482  
QY 1211 NTPQTTLSQVQ--TSSSPSTNTFIASYT---DGSII 1243  
Db 1483 PSAITTSVEHTESTISITTKVPVTVTSSDTNGELI 1520  
RESULT 13  
ID YM96\_YEAST STANDARD; PRG; 1140 AA.  
AC Q04893;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE Hypothetical 113.1 kDa protein in PRE5-FET4 intergenic region.  
GN YMR317W OR YMR924.09  
OS Saccharomyces cerevisiae (Baker's yeast)  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_taxid=4932.  
RN [1] J  
RP SEQUENCE FROM N.A.  
RC STRAIN=S288C / AB972;  
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;  
Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
RL -I- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; Z54141; CAAB0835.1; -  
DR SGD; S0004936; YMR317W.  
KW Hypothetical protein; Repeat.  
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;  
Query Match 8.0%; Score 521.5; DB 1; Length 1140;  
Best Local Similarity 24.1%; Pred. No. 8.8e-16;  
Matches 243; Conservative 193; Mismatches 379; Indels 195; Gaps 39;  
QY 320 AGSNGIVIVATRR---TVTDSSTAVTTLFPNPSVDKTKIEILQIPITTTTTSYGVTT 376  
Db 4 SCKSTATTSHSSTTTTSTSTTT-----PTTSTTS---TT 40  
QY 377 SYLTATPAGTATVIVDPVHTTTVTSSEMTGTTTTTRTNPDTSDTVVQVPLP 436  
Db 41 STKVTTSP-----EIVSSSSTLVSSVVPETSSSL-----SDTASILSESLS- 88  
QY 437 TVSTTEYWSQSFATTT-----TVTAPCGTDTVIREPPNHTVTTTEYWSQSFATTTVT 491  
Db 89 IESSLVSSTSDISSTVNDVESSTGSPNSVSALSSTNAQLSSSTTE--TDSISSAIO 146  
QY 492 APP-----GGTDS-----VIIREPPNPTVTTTEYWSQSFATTTTVPAG--GGTDSVI 537  
Db 147 SSPQTSNGGSGSSSEPLGKSVLETTASSDSTAVTSTFTTLTDVSSSPKSSSGS 206  
QY 538 IREPPNPTVTTTEYWSQSFATTTTVPAGTGDVIREPPNHTVTTTEYWSQSFATTT 595  
Db 207 -----TSVGTSTDSARKEVFSSTSDVSSLSSTSS-----PASSTISLTPES--SRI 252  
QY 596 TTVTAPCGTDTVIREPPNHTVTTTEYWSQSFATTTTVPAGTGDVIREPPNPTVT 655  
Db 253 LSITSSP-----VSSEAPSATSSSVSSSEASSTSSVSSEAPLATSSVSSSEAPS----- 302

```
QY 656 TTEYSQSYATTTTITAPPGETDVIIRPPNHTVTTEYSQSYATTTTITAPPGETDT 715
DB 303 -----STSSVSSSEAPSTSSVSSEISSTTSSVSSEAPLATSS 342
QY 716 VLIREPPNHTVTTEYSQSYATTTTITAPPGETDVIIRPPNHTVTTEYSQSPAT 775
DB 343 VVSSEAPSTSSVSSEISSTTSSVSSEAPLATSSVSSEAPSTSSVSSEAPSTSS 402
QY 776 TTVTAPPGTDTVIIESMSSKLTSSNDITSIIPSPRPHVNSTTSDLTSSSSNN 835
DB 403 SVSSEAPSTSSVSSEISSTTSSVSSEAPLATSSVSSEAPSTSSVSSEAPSTSS 446
QY 836 TPTSSDGMLLSTTLTSEITTEL-ICSDGKESRLSSSGVITNPDSN--ESSIV 891
DB 447 AISSLASRLFSKNTSVTLVATEASSVTSRLPSSETLASNLSLSTGYNSTVS 506
QY 892 TSTVPTASTMSDLSLSTGISATS-----SDNVKSGVSVTTTSTVTR-----IQTPNP 941
DB 507 TTTSAASTLGSVSSSRNATSKTSSSTSDLSKSSVIEGNSVTTTSPASISLTSAP 566
QY 942 LSSSVTLTQSLSIPIVSESESKVFTTSCNDQSGTHDSOSTSTEIE-IYITSTKVLPP 1000
DB 567 LPSVMSDITS-----SEASS-----ISSNLASSAPSDNNSTIASASLIVTKNSVVS 616
QY 1001 VVSSNLTSEPTNREPOTTLTSTNSITDITTSOPTGDNQDNTSTNVPVATSTL 1060
DB 617 IVSSIT--SSETTN-----ESNLATSTSLLSNKATARSLSTS--NATASKNVPTGTFSM 668
QY 1061 ASASEEDNKGSHESASTSLKPSNGENS--GLTITSTEIATTTT--PTEAPSPVSSGT 1115
DB 669 SSTSHTVPTGFTSSASALAINSTVSSSLAGVSFTSPESPSTTLTSTVSEAPS--TVSSMT 727
QY 1116 DVTTPEP-----TDTRPOTTLTSTKTNSELVATTQATNENGGKSPSLDTSLLTQTS 1170
DB 728 --TSAPFINNSTSARPSSTASFISTSSISSVPLASGDVTSLSLAHNLTFSPAPSTSS 785
QY 1171 -----STANSSELVT-----SGSVTGA-----VASASNDQSHSTSVTNSNS 1207
DB 786 AOLVSKSTSSSLTVPRIDRSGNSSTASRIATSLPNKTTFVSSLSSTSAHARNFTSV 845
QY 1208 IVSNPTOTLLSQOVTSSTPNTFTASTYDVGSGTIOHSTWLYGLITLSS 1257
DB 846 LATAQIETLTSTVNCSPNTPNNTKTV-----IVSRETATGTVTS 890
```

## RESULT 14

```
QY_K82_YEAST STANDARD; PRT: 1169 AA.
P36170;
DT 01-JUN-1994 (Rel. 29, Created)
DE 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-JUN-1994 (Rel. 29, Last annotation update)
DE Hypothetical 122.2 kDa protein in SIR1.3 region precursor.
GN YKRL02W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Gallion L., DuJon B.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: CELL WALL (BY SIMILARITY).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY).
CC -1- SIMILARITY: STRONG, TO YEAST PROTEIN PLOI.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

```
CC or send an email to license@isb-sib.ch.
CC EMBL: 228327; CAAB2182.1; -----
DR PIR: S38181; S38181
DR SGD: S0001810; YKRL02W.
KW Hypothetical protein: Glycoprotein; Repeat; Cell wall; Signal.
FT SIGNAL 27 1169
FT CHAIN 27 1169
FT DOMAIN 274 609
FT CARBOHYD 122 122
FT CARBOHYD 157 157
FT CARBOHYD 270 279
FT CARBOHYD 389 389
FT CARBOHYD 452 452
FT CARBOHYD 515 515
FT CARBOHYD 578 578
FT CARBOHYD 658 658
FT CARBOHYD 686 686
FT CARBOHYD 809 809
FT CARBOHYD 1092 1092
FT CARBOHYD 1099 1099
FT CARBOHYD 1169 1169
SQ SEQUENCE 1169 AA; 122164 MW; 940C6DFB3569C669 CRC64;
```

Query Match 7.7%; Score 501; DB 1; Length 1169;  
Best Local Similarity 23.1%; Pred. No. 6.9e-15;  
Matches 313; Conservative 208; Mismatches 528; Indels 308; Gaps 53;

```
QY 6 TLLFLYLSTASAKTIT-----GVFDSFNSLTWNSAANYAFKPGYPTWNAVLG 53
DB 11 TGLEL-LSVANVALGTTTEACLPAGEKKGWMTINFYQSLKDSSTYS--NPSYMAGYADA 67
QY 54 WSLDGTGANPGDFTFLNNPCVFKYTTQSVDLTADGVKATQCFYSGE-----EFTT 106
DB 68 EKLGSVSGQTKLSIDYSIPCGASDTCACSD--DATEYSASQVVPVKRGVKLCSDNTT 124
QY 107 FSTLT-----CTVNDALKSKAFGTVTLPIAFNVGGTSDTLEDSKCFAGTNTVT 159
DB 125 LSSKTEKRENDCCOGAAWSS-DLFGFTVTTNTVEMTG-----YFLPPKGTYT 175
QY 160 FNDGDKSIDVEFEKSDVPSAVLYASRVMSPLNKVTLFVAPOCENGYSGTMGFSSS 219
DB 176 FG-----PATVDSAIL-----SV 189
QY 220 NGDVAIDCSNIH---IGIT---KGLDNH--YPVSSSEFSTKCTCTNSGLOIKYON-VP 269
DB 190 GGNVAFECCKEQEPITSTDFTINGIKPNADAPTDIKGSTYMYAGYYPPIKIVSNV 249
QY 270 AGYRPFDAISATDVNQ---YTLAYTD-----YTCAGSRLOSKPFTL 310
DB 250 WGTLPVSVVLPDGTENVDDFEGVYFSDDNATQAHCSVPNPAEHARTCVSATSS----- 304
QY 311 RWGYNKSDAGSNVIVATRTVDTSTAVTTLFPNPSVDKTKTTEILQIPITTTITS 370
DB 305 -WSSSE-----VCTECTETSTSYTPVTVSSSSSEVCTECTETSTSTP 352
QY 371 YVGVTSTLYLTK-APIGETATVIVDPHYHTTTTVMETGTTITTT---TRNPTDSIDT 426
DB 353 YVTSSSSSSEVCTECTETSTSYTPVTVSSSSAAANYTSFSSSEVCTECTETSTST 412
QY 427 VVQVPLPNPTVTTEYSQSFATTTTATPGGTDVTVIIRPPNHTVTTEYSQSFAT 486
DB 413 STPVV-----TSSSSSEVCTECTE-----TESTSYTPVTVSSSSAAANYTSFSS 459
QY 487 TTTVATPGGTDVTVIIRPPNPTVTTEYSQSFATTTTATPGGTDVTVIIRPPNPTV 546
DB 460 SEVCTECTETES-----TSTSTPVTVSSSSSEVCTECTETSTSYTPVTVSSSS 510
QY 547 TTEYSQSYATTTTITAPPGETDVIIRPPNHTVTTEYSQSYATTTTITAPPGETD 606
DB 511 TAAANYTSFSSSSSEVCTECTETES-----TSTSTPVTVSSSSSEVCTECTETET 561
QY 607 TVIIRPPNHTVTTEYSQSFATTTTATPGGTDVTVIIRPPNPTVTTEYSQSYAT 666
```

Db 562 STSYVTPVSSAAAYTSFSSSEVCTECTETES-----ISTSPATSTGT 612  
 Qy 667 TTTTAPGCTDVLREPNNHTVTTTYSQSYATTTTAPGCTDVLREPNNHTV 726  
 Db 613 ATSTASTNTWLSLVQ-----TDTTYSLSSTVSEHTNAP-----TSVSNASTFI 661  
 Qy 727 TTTTYSQSYATTTTAPGCTDVLREPNNHTV-----TTEYSQSYATTTTAP 781  
 Db 662 SSNGSKSVYTSIHSITP-----MTPSNQVTSVVSVTPTTSESSESASVTL 713  
 Qy 782 PGCTDVLIIYESSMSSKISTSNITSPF---SRPHYVNSTSDLSFESSMNTPT 838  
 Db 714 P-STTSEFPKSTMTKVVYSSISSTPLNLSYDTTSKDTGVSSTSSVLSISLPSY 772  
 Qy 839 STSSD---GMLLSSTLVTSEITTELCSDGKESRLSSSSGIVNPDNNESSIVTST 894  
 Db 773 SASSEQIFHSISVSSNGQALTSFSTKVSSESESHRTSPIT-----SSSEGIKSSG 825  
 Qy 895 VPTASTWSDLSSTDGISATSDNYSKSGVSTTETSVTTTQTPNPLSSSVTLQLSS 954  
 Db 826 VRIESTSTSFSPHE--TSTASTSVQISSOFVTPSPISVTA---PRSTGLNSQTE--- 876  
 Qy 955 IPSVSESESVKVTSTNGDNQSGTHDSQSTSTEIEIVTTSKVLPPVVSNTD----- 1007  
 Db 877 ----STNSSKETMSSE-NSASVMPSSSATSPKTKVTSDET-----SSGFRDRTTV 924  
 Qy 1008 -LTSPNTREQPTTLTSTT-----SNSITEDITTSOPTGNDGNTSTNPNVPTVATST 1059  
 Db 925 RMTSETSTPNEQTLITVSSCESNCSNTVSSAVSTATTINGITTEYTWCPLSATE- 983  
 Qy 1060 LASASEEDNKGSHESASTKPKSGENSGLTSTETATTSPEAP---SPAVSSGTD 1116  
 Db 984 LTTVSKLESE-----EKTLLIIVTSCSGVCSSETASPAIVSTATATVND 1027  
 Qy 1117 VTTEPTDTRQPTT-LSTTSKTNSEL-----VATTOATNENGKSPSTDLTSLT-T 1166  
 Db 1028 VTVYVTSWSPQATNKLVSSDIDENSASKASFVSEAEATKSISRNNNFVPTSGTSTETHT 1087  
 Qy 1167 GTSASTSANELVTSVGVGAVASANDOSHSTVYNSNST-VSNTPORTLSQOYVTS 1225  
 Db 1088 YTTNASENDNV-----SASEAVS-SKSVTNVPLISVQQPRTGPASSMIGS- 1134  
 Qy 1226 PSTNTFTASTYDGSIGIOHSTWLYGLITL--LSLEI 1260  
 Db 1135 -STASLEWSSFLG---IANH-----LUTNSGISIFI 1161

RESULT 15  
 50\_HSV11 STANDARD: PRT: 670 AA.  
 AC Q00130;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Hypothetical gene 50 protein.  
 GN 50.  
 OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC unclassified Herpesviridae.  
 OX NCBI\_TaxID=10401;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AUBURN 1;  
 RX MEDLINE=92087450; PubMed=1727613;  
 RA Davison A.J.;  
 RT "Channel catfish virus: a new type of herpesvirus.";  
 RL Virology 186:9-14(1992).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR ENBL: W75136; AAA88153.1; -  
 DR PIR: F36791; F36791.  
 KW Hypothetical protein; Repeat.  
 FT REPEAT 143 158  
 FT REPEAT 171 186  
 FT REPEAT 200 214  
 FT REPEAT 215 233  
 FT REPEAT 234 252  
 FT REPEAT 253 268  
 FT REPEAT 279 293  
 FT REPEAT 294 309  
 FT REPEAT 320 334  
 FT REPEAT 335 349  
 FT REPEAT 362 376  
 FT REPEAT 377 391  
 FT REPEAT 392 406  
 FT REPEAT 407 421  
 FT REPEAT 422 436  
 FT REPEAT 437 452  
 FT REPEAT 464 477  
 FT REPEAT 478 493  
 FT REPEAT 504 517  
 FT REPEAT 518 531  
 FT REPEAT 532 545  
 FT REPEAT 546 559  
 FT REPEAT 560 573  
 FT REPEAT 574 587  
 FT REPEAT 588 601  
 FT REPEAT 602 615  
 FT REPEAT 616 629  
 FT REPEAT 629 644  
 SQ SEQUENCE 670 AA; 64174 MW; 2864A781C519EBB4 CRC64;  
 Query Match 7.28; Score 464.5; DB 1; Length 670;  
 Best Local Similarity 24.38; Pred No 1.4e-13;  
 Matches 215; Conservative 71; Mismatches 294; Indels 303; Gaps 34;  
 Qy 297 TCAGSRLOSQKFTLRWT--GYKNS--DAGSNGIVVATFTVTDST-----TAV 341  
 Db 9 TCGMSLLASNLALGVVRSYNETPTGPELETITTTVARMTDTAKPDPFGDAVTGT 68  
 Qy 342 TTLPNPNSVDKTIETILOPIPTTTTITSYGVVTSYLTKTAP---IGETATVI--VDVP 396  
 Db 69 QPVPREPS-----SLRPTPSLAHTTISKMISLGTIRPTIPGVTTIPNTDAP 117  
 Qy 397 -----YHTTTVTSEWGTITTTTRTNPTSDIDTVVQVPLNPVTSTTTEYWSQSFA 451  
 Db 118 VDPGSVHTTARVVDIT-TKOTPTTPATPAGANDTANITATPAGANDT-----ANI 168  
 Qy 452 TTTTAPPGTDTVIIREPPNHTVTTTTEYWSQSFAATTTTTPAPPGTDSVIIREPPNPTVT 511  
 Db 169 TTAT-PAGANDTANI-----TTATPAGANDTAVTTTATPAGANDT-----208  
 Qy 512 TTEYWSQSFAATTTTTPAPPGTDSVIIREPPNPTVTTTEYWSQSFAATTTTTPAPPGTDS 571  
 Db 209 -----AVTTTTPAGANDTA-----NDTV-----VTTTPAMPAGANDT 242  
 Qy 572 VIIREPPNHTVTTTTEYWSQSFAATTTTTPAPPGTDTVIIREPPNHTVTTTTEYWSQSFA 631  
 Db 243 -----ANGTAVTTTPAMPAGANDTANIT-----TATPT-----GANDT 275  
 Qy 632 TTTTGPSTGDTVIIREPPNPT-----VTTTTEYWSQSFA-----TTTITAPPGTDS 678  
 Db 276 ANVTMPAGANDTAVTTTTPAMPAGANDTANIT-----ATPAGANDTANVTMPAGANDT 327  
 Qy 679 TVLIREPPNHTVTTTTEYWSQSFAATTTTTPAPPGTDTVIIREPPNHTVTTTTEYWSQSFA 738  
 Db 328 TV-----VTTTTPAMPAGANDTANVTMPAGANDT-----357

[illegible]

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 15:32:19 ; Search time 46.21 Seconds  
(without alignments) 4717.023 Million cell updates/sec

Title: US-09-715-876-8  
 Perfect score: 6495  
 Sequence: 1 MLQQFTLLFLVLTASAKTI.....SIQHSWYGLITLLSLFI 1360  
 Scoring table: BLOSUM62

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

```

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```

database :
SPTRMBL_19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mouse:*
8: sp.organelle:*
9: sp.phase:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.virus:*
14: sp.virus:*
15: sp.unclassified:*
16: sp.bacteriap:*
17: sp.archaeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Query			ID	Description	
	Match	Score	Length			
1	62.8	1078.5	1047	3	O93865 candida alb	
2	57.2	2716.5	1523	3	O9HF44 candida alb	
3	43.2	2793.5	1443	3	O9YF82 candida alb	
4	39.7	2576.5	2209	3	O9Y743 candida alb	
5	30.6	1986.5	468	3	O9HGK6 candida alb	
6	28.1	1823	281	469	3	O9UR90 candida alb
7	27.4	1780.5	27.4	468	3	O9C471 candida alb
8	22.6	1471	22.6	336	3	O9HF70 candida alb
9	18.3	1191	18.3	338	3	O9HF72 candida alb
10	18.1	1276	18.1	331	3	O9HF69 candida alb
11	15.4	996.5	433	3	O9YF81 candida alb	
12	15.1	868.5	353	3	O9HF71 candida alb	
13	12.7	831.5	1392	3	O96W08 candida alb	
14	12.7	824	1272	3	O9HG16 schizosac	
15	10.8	698.5	1079	5	O9N457 candida alb	
16					O9n457 caenorhabdi	

17	685	10.5	2586	5	Q9VTK8	Q9vtk8 drosophila
18	674.5	10.4	1220	3	Q9C0Y2	Q9c0y2 schizosacch
19	649.5	10.0	1275	5	Q76602	Q76602 caenorhabd1
20	646.5	10.0	2231	5	Q91365	Q91365 caenorhabd1
21	637.5	9.8	2272	16	Q93944	Q93944 staphylococ
22	632.5	9.7	3570	4	Q99552	Q99552 homo sapien
23	620	9.5	3072	2	Q939N5	Q939n5 streptococ
24	601	9.3	1795	5	Q76894	Q76894 drosophila
25	600.5	9.2	2178	2	Q9KWR3	Q9kwr3 schizosacch
26	594.5	9.2	3971	3	Q96WV6	Q96wv6 schizosacch
27	582.5	9.0	1752	2	Q9AE52	Q9ae52 ruminococcu
28	577.8	8.9	1283	3	Q9USQ3	Q9usq3 schizosacch
29	577.5	8.9	957	4	Q14651	Q14651 schizosacch
30	577.5	8.9	1217	4	Q9UKW9	Q9ukw9 homo sapien
31	574	8.8	1283	3	Q9URU4	Q9uru4 schizosacch
32	570.5	8.8	1858	5	Q9C470	Q9c470 candida alb
33	553.5	8.5	1459	5	Q17084	Q17084 caenorhabd1
34	553.5	8.5	3178	5	Q969D4	Q969d4 caenorhabd1
35	545.5	8.4	957	4	Q9UKN0	Q9ukn0 homo sapien
36	538.5	8.3	406	3	Q9URP9	Q9urp9 candida alb
37	538.5	8.3	1203	5	Q9NSK0	Q9nsk0 schizosacch
38	534	8.2	849	3	P87107	P87107 saccharomyc
39	533	8.2	407	3	Q74661	Q74661 candida alb
40	531.5	8.2	407	3	Q74658	Q74658 candida alb
41	520	8.1	407	3	Q74659	Q74659 candida alb
42	526	8.1	1229	5	Q94185	Q94185 caenorhabd1
43	515	7.9	4776	16	Q97P71	Q97p71 streptococ
44	511.5	7.8	3507	5	Q23587	Q23587 caenorhabd1
45	503.5	7.8	1844	5	Q23579	Q23579 caenorhabd1

## ALIGNMENTS

181	QY	181	SAYLAVASRMP	SNKVTLLFVAPOCENG	YSGTGMGSSNGDVA	ICDSNIHGITKGLND	240
181	Db	181	KGVLTD	SRVIRP	SNKSVTLFVAPOCANG	YSGTGMGSSNGDVA	ICDSNIHVGITKGLND
241	QY	241	WNYPVSE	SFSYTKTCSNG	IOIKYONPAGYR	REFDAYISATD	NOYNYTLAYINDYTCAG
241	Db	241	WNYPVSE	SFSYTKTCSNG	IFITKKNPAGYR	REFDAYISATD	VNSYTLSEANEYTCAG
301	QY	301	SRLOSKP	FLRW	GYKNSDAGSNG	IVIVATRTV	TOSTAVTTLPENPSVDKTKTEILQ
301	Db	301	GYMORAP	FLRW	GYRNSDAGSNG	IVIVATRTV	TOSTAVTTLPEFDRNDRKTKTEILK
361	QY	361	PIPTATIT	SYGVGTV	SYLTKTAP	TGSEATV	VDVPHYTTVTSEMTGTITTTTTRTP
361	Db	361	PIPTATIT	SYGVGTV	SYLSIKTAP	TGSEATV	VDVPHYTTVTSKWTGCTITSYTHTRP
421	QY	421	TDSDT	VVVVQV	LPNPTVST	TEYWSQSFAT	TTVTAPPGTDTVIIRPPPHNVTTEYX
421	Db	421	TDSDT	VIVQV	SPNPVT	TTTEYWSQSEAT	TTITGPGNTDVLIRPPPHNVTTEYX
481	QY	481	SQSAT	FATTTV	TAPPGCTSV	LIIRPPNPVT	TTTEYWSQSFATTTVTAPPGTDSVIIRE
481	Db	481	SESYT	ISTFTAP	PGCTDSV	IIKEPPNPVT	TTTEYWSQSEYTTSTFTAPPGTDSVIRKE
541	QY	541	PNPNVT	TTTEYWSQ	SVATTTVT	TAPPGCTDSV	LIIRPPNPVT
541	Db	541	PNPNVT	TTTEYWSQ	SVATTTVT	TAPPGCTDVL	IRPPNPVT
601	QY	601	PPGCTD	TV	LIIRPPNPVT	TTTEYWSQ	FATTTVTGPPSGTDTVIIRPPNPVTTEYX
601	Db	601	PPGCTD	TV	LIIRPPNH	-----	-----TVTTEYX
661	QY	661	SQSYAT	TTTITAP	PGCTDVL	IRPPNHVT	TTTEYWSQYATTTVTAPPGCTDVLIRE
625	Db	625	SQSYAT	TTTITAP	PGCTDVL	IRPPNHVT	TTTEYWSQYATTTITAPPGCTDVLIRE
721	QY	721	PNNHVT	TTTEYWSQ	SVATTTVT	TAPPGCTDVL	IRPPNPVT
685	Db	685	PNNHVT	TTTEYWSQ	SVATTTVT	TAPPGCTDVL	IRPPNPVT
781	QY	781	PPGCTD	TV	LIIRPPNPVT	TTTIPSRR	PHYVNSTTSDLTSESSMNTPTS
745	Db	745	PPGCTD	TV	LIIDTMS	SEIS	-----SFRPHYVNHVT
841	QY	841	SSDGM	LLSTLTV	SEETETEL	ICSGKRC	SLSSGIVTNPDSN
777	Db	777	-----	-----	-----	-----	-----
901	Db	901	MSD	SLSDG	ISATSD	NKVSQSV	TTTETS-VTITOTNP
831	Db	831	ESTS-QSPG	-----	-----	-----	-----
960	QY	960	ESEK	SVTFT	NGDNQSG	THDSQ	TSSTIE
866	Db	866	TTSE	VEFTTK	GNNGNG	PEFSH	-----
1020	QY	1020	TTLS	TTSTNS	ITDITTS	QPTGDC	NDTSSNPNVPT
903	Db	903	-----	-----	-----	-----	-----
1080	QY	1080	LKPS	MG	ENGLT	TSBIE	-----
909	Db	909	-----	-----	-----	-----	-----
1137	QY	1137	TNSEL	VAT	TQNTEN	-----	-----
947	Db	947	STSV	-----	-----	-----	-----
1196	QY	1196	QSHV	STSV	NSNS	IVNT	POTLLSQOVTSS
1000	Db	1000	QSHV	STSV	NSNS	-----	-----
1255	QY	1255	QSHV	STSV	NSNS	IVNT	POTLLSQOVTSS
1042	Db	1042	QSHV	STSV	NSNS	-----	-----

```

Db 601 PPGGDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 560
Qy 661 SQSATTTTITAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 706
Db 661 SQSATTTTITAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 720
Qy 707 PPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 744
Db 721 PPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 780
Qy 745 PPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 800
Db 781 PPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 840
Qy 801 TSSNDITSIIP-----FSRPHYNSTSDLSFSSSMNTPTIS-----SDGWL 846
Db 841 SQSATTTTITAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 897
Qy 847 L-----SSTLVTESETTELI-----CSDGKCSRLSSSGI 879
Db 898 IREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 957
Qy 880 VTNPDSNESSIV-----TSVPTASTMSDLSSTDGISA-----TSS 916
Db 958 VTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 1017
Qy 917 DNKSGKSVTTEY-----SVTT----- 934
Db 1018 EYWSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 1077
Qy 935 IOTPNPL-----SSSVTLSTLSSIPSVSE-----SES-----K 964
Db 1078 IREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 1137
Qy 965 VTFTSGN-----QSGTHDSQST-----EIEVTSTKVLPPVSSNT- 1006
Db 1138 ITAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 1194
Qy 1007 ---DLTSEPTREOPTLSTNSNITE-----DLTTSOPTGDNQNTS----- 1047
Db 1195 HLPSSSKPVDIPSDVVPSTNDLSLSTGSENGETSVAISTTFCDDENGCSOTSIPOGS 1254
Qy 1048 -STNPVPVATST-LASASEEDNKGSHESAST-----SLK-----PSMG 1085
Db 1255 VRTAPTATPTPIIGDNGSGKSGELSGTGVTTNTATPDVPSTKVPSPGAPCTG 1314
Qy 1086 ENSGLTTEIEATTSPTAPSPAVSGSDTDTPTDT-----REOPTL 1131
Db 1315 VPPPLAPSTETQTNVPGSPNIPATGT-TDIIRESTTVSHVTGNCTGVPMNPNALT 1373
Qy 1132 STTSKTNSELVATTQATNE-----NGKSPSTDLTS--SLTGTGSASTSANSSELVTSGS 1183
Db 1374 TSTSLTG-----ATNSATNPSHETGVNTRGSGTNIIVPPSSATATVIGTDNGATTKGQ 1429
Qy 1184 VTGGAVAS-----ASNDQSHSTSVTNSNSVNTQTTLTSLQVTS 1224
Db 1430 DTAGNSNGFATTNQTGGNPNQPCNTGTGEPVGTGQSVSTSPITLSQVTS 1489
Qy 1225 SPSTNFIASITDGSIOHSTWGLTLLSLF 1259
Db 1490 LISTP--LASTFDGSGSIQVHSGMLVLLTAISIF 1522

RESULT 3
Qy8F2 ID QY8F2 PRELIMINARY; PRT; 1443 AA.
AC QY8F2;
DT 01-NOV-1999 (TEMBLrel. 12, Created)
DT 01-NOV-1999 (TEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE AGGLUTININ-LIKE PROTEIN 6.
GN ALS6.

```

```

OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RC MEDLINE=20321177; PubMed=10861907;
RA Hoyer L.L., Hecht J.E.;
RA Ytter ALS6 and ALS7 genes of Candida albicans.;
RL Yeast 16:847-855(2000).
DR EMBL; AF075293; AAD42033.1; -.
DR IncerPro; IP0003015; HLH_MYC.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
SQ SEQUENCE 1443 AA; 151476 MW; DCBD693F4C435809 CRC64;

Query Match 43.0%; Score 2791.5; DB 3; Length 1443;
Best Local Similarity 44.2%; Pred. No. 6.7e-121;
Matches 642; Conservative 202; Mismatches 387; Indels 223; Gaps 35;

Qy 8 LFYLSIASAKTITGVDFSNLSLWSNAANAFKPGCYPTMNAVGLMSLDGTSANPGDTF 67
Db 9 LFYCTIAMAKTISGVTFSNLSLTYTNGNYPYGGPYPTWAVLGNLSDGTLASPGDTF 68
Qy 68 TLMNCPVKYTSQTSVDLTADGVKATCFYSGEEFTTFTSLACTVNDALKSSIKAFGT 127
Db 69 TLMVPCVKFTTQTSVDLTANGVKYATCTFHAGEDTFTFSMSCVYVNGLSNIRAFGT 128
Qy 128 VTLPTAFNVTGTSSTDLSDSKCFTAGTNTVTFNDGDKDISIDVEFEKSTVDPFSAYLYAS 187
Db 129 VRLPTSFNVTGTSVNIQDSKCFCTAGTNTVTFDGDHKTSTVNFPKTPQSSSLVYFA 188
Qy 188 RVNPSLVNVTFLVAPQCNVTGTMGFSNGDVAIDCSNIHIGITGLNDWNPVSS 247
Db 189 RVPLSLDLKSLVWSQCTAGYASGVLFSGATKDDVTIDCSTIHVGTNGLSNMNPVSS 248
Qy 248 EFSYTKCTGNGIQKQNPAGYRPFIDAYI--SATDVNQYTLAYTNDYTCAGSRLOS 305
Db 249 EFSYTKCTGNSITITENVPAGYRPFIDSVYKKSATATNGFNLTNYNCMDGKGN 308
Qy 306 KPFTLWTKYKNSDAGSNIIVATRTVTDSTAVTTLFPNPSVDKTKTIEILOPIPT 365
Db 309 DPLTFWTSYNSDAGSNGAAVVTTRVTDSTAITLTPDPTVDKTKTIEVIEPIPT 368
Qy 366 TITSYGVVTSYLTATDGTATVIVDPVHTTTVTSEWGTITTTTNTPTSD 425
Db 369 TITSYGISTSLSTKATIGTATVVDVPHHTTTITSIWTSATSTSTVNTPTSD 428
Qy 426 TVVQVPLNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 485
Db 429 TVVQVPSNPVTTFQFWSGSVPTTETVTGQGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 488
Qy 486 TTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 545
Db 489 TTTTNNPPTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 548
Qy 546 VTTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 605
Db 549 VTTTEFSQSFATTTTAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 608
Qy 606 DTVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 649
Db 609 DSVIREPPNPTVTTEFSQSFATTTTAPPGTDSVIREPPNPTVTTEW 668
Qy 650 -----PNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 669
Db 669 ISSAQESSSSVEQSLSADETSSIVELSSRDIPSSSISGLTSSSTVSSVSSSSTS 728
Qy 670 ITAPPGTDSVIREPPNPTVTTEYSQSFATTTTAPPGTDSVIREPPNPTVTTEW 714
Db 729 ESSIASYDYSISSSIESSTLSLSDRCSSISDITTSFWDSSSDLESTSTWSSSIDAOS 788

```

QY 715 TVLIREPNNHTVTTEWWSQSYATTTTVPAGGDTVTIIRP-----PNTV 762  
Db 789 SHLVGSVNSISFQELSSSSSEESTF-----ATDALVSSDASSILSDSTSSYFSPSTI 843  
QY 763 TWTTEW-----SOSFATT-----TWTAPGGT-DTVIYESMSSSKISVS 802  
Db 844 SSSDDFPHTIAGESDSLSLFSITSVELSSDSVSLTSPASSFSSSLNSDSSSPSSD 903  
QY 803 SNDI-----TSLIPFS-----RHYVNSTSDLTSESSMNTPTSISSDGM 845  
Db 904 QSDILTSSTFLVVPFSLSSSSLSLTYPHYVNSTVHYASESSSVASPSMASSEAN 963  
QY 846 LLSSTLVTESEFTTELICSDGKE---CSRLLSSSSGIVTN-PDSNESSIVTSTVPTASTM 901  
Db 964 --DDIYLTSESTDTISGTSSTVFCRRDNGDGLVTCMPSSSIDSEQTSDVTTTSS 1021  
QY 902 SDSLSTDG-ISATSSDNV--SKSGVSVTTSVTTIOTTPNPLSSSVTSLSLQSSIPSV 958  
Db 1022 VASSPTSAEQSIITDNPIDSSQTSASSSTKLSVSDTVVNSISLSETS--TLSSDDST 1079  
QY 959 SESEKVTFTNGD---NCGTHDSOSTSTEIRIVTTS---STKVLPPVYSSNLTLS 1011  
Db 1080 S-SDISISSTNSDTGNVAGSSHTSTASIKESSIQKTGVTLLSSSVLSTKLSSTDI 1138  
QY 1012 PNTREOPTLTSTNSITD---LITTSOP-----TGDNQ-----DNTSNTNPVT 1054  
Db 1139 LITTELTITTELTITIEDNEPNTFTTSPSSHSEIFSSSDNSVLSKQVDSSTVEIPPV 1198  
QY 1055 VATSLASAEEDKNGSHESASTSLKPSMGENS-----GLTSTEIEATTTSPT 1106  
Db 1199 TDTTVSSVSHST---ASTA---TLGENSEFKVASAPVNTETLSRSTSSSNHAT 1249  
QY 1107 -----PSPVSSCT--DVTEPTDREOPTTLSTTS-KTNSBELVATTQAT 1148  
Db 1250 ESSGTVKSEASAEATPPTSTONRLSYSTEBAKSGTYPNSGNTNLMTESQVAAPTST 1309  
QY 1149 N---ENGKSPSTDLTSLTGTSGASTSANSBLVTS---GSVTGAVASANDQS--- 1197  
Db 1310 SVLTANPVVTSFDDKSSAAVNQPSKTSIEESIGSLDSVNETNNGFIATLSQSEAPNSL 1369  
QY 1198 -HSTSVT-----NSNSIVSNTPOTLISQOVTSSSPSTNTASTYDGSGLIQH 1245  
Db 1370 IHSEISITMAKTDDASINGDAAANSOPTTLIOQVATS--SYNOPLITYAGSSATKH 1427  
QY 1246 STWLYGLITLLSLF 1259  
Db 1428 PSMLLKAFISVALFF 1441

FULT 4  
ID Q9V743 PRELIMINARY; PRT; 1270 AA.  
AC Q9V743;  
DT 01-NOV-1999 (TREMBLrel. 12, Created)  
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE AGGLUTININ-LIKE PROTEIN.  
GN ALS5.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RX MEDLINE=21064501; PubMed=111124701;  
RA Hoyer L.L., Hecht J.E.;  
RT "The ALS5 gene of Candida albicans and analysis of the Als5p N-  
RT terminal domain".  
RL Yeast 18:49-60(2001).  
DR EMBL: AF068866; AAD32849.1; .  
SQ SEQUENCE 1270 AA; 133337 MW; D8E5FA5853F6D5C5 CRC64;

Query Match 41.2%; Score 2673.5; DB 3; Length 1270;  
Best Local Similarity 45.9%; Pred. No. 1.6e-115;  
Matches 629; Conservative 187; Mismatches 340; Indels 215; Caps 39;  
QY 1 MLQOFTLLFLXLSIASAKTITGVDFSNLSWNSAANYAFKPGYPTWNAVGLWSLQDTS 60  
Db 1 MIOQFTLLFLXLSFATAKATIGTFNSIDSLTWSNAGNYAFKPGYPTWNAVGLWSLQDTS 60  
QY 61 ANPGDFTLLNMPCKVFTTSQTSVLTADGVKATPCFYSGEEFTTFTLTCTVNDALKS 120  
Db 61 ANPGDFTLLNMPCKVFTTSQTSVLTADGVKATPCFYSGEEFTTFTLTCTVNDALKS 120  
QY 121 SIKAFGTVLPFAFNVGGTSSDLEDKSCFTAGNTVTFNDGKDIDISIDVEFEKSVDP 180  
Db 121 SIKAFGTVLPFAFNVGGTSSDLEDKSCFTAGNTVTFNDGKDIDISIDVEFEKSVDR 180  
QY 181 SAYLASRYMPSLNKVTTLFVAPQCENGYTSCTMGFSSNGDVAIDCSNIHIGITKGLND 240  
Db 181 SGYLTTSRMPNLKIAFLYVAPQCENGYTSCTMGFSTYGDVAIDCSNHLIGISGVND 240  
QY 241 WNPVSESFSTKCTCTSNIGIQKYNVPAGYRPIDAYISATDVNQYTLATNDYTCAG 300  
Db 241 WNPVSESFSTKCTCTSNIGIQKYNVPAGYRPIDAYISATDVNQYTLATNDYTCVD 300  
QY 301 SRLQKPFTRWTKYKNSDAGSNGIVATVTRVTDSTAVTTLPPNPSVDKTKIEIQ 360  
Db 301 DYWOHAPFTLKWTGYKNSDAGSNGIVATVTRVTDSTAVTTLPPNPSVDKTKIEIQ 360  
QY 361 PIPTTTTTSYGVVTSYLTAPIGETATVIVDVYHTTTSVETGTTTITTRNP 420  
Db 361 PIPTTTTTSYGVVTSYLTAPIGETATVIVDVYHTTTSVETGTTTITTRNP 420  
QY 421 TDSIDTVVQVPLNPVTTEWWSQSYATTTTVPAGGDTVTIIRPNNHTVTTEWWS 480  
Db 421 TDSIDTVVQVPLNPVTTEWWSQSYATTTTVPAGGDTVTIIRPNNHTVTTEWWS 480  
QY 481 SQSFATTTVTPAGGDTVTIIRPNNHTVTTEWWSQSYATTTTVPAGGDTVTIIRP 527  
Db 481 SESATFTTITGPGTSDVIHDLPESSSTIATLESSDNISSAGESSSVESQSSIV 540  
QY 528 APGCGTSVIREPNNHTVTTEWWSQSYATTTTVPAGGDTVTIIRPNNHTVTTEWWS 587  
Db 541 GLSSSDIPLUSSDMFS-----SSSGLTSSESTVSVSDSSSSSELSTFS 587  
QY 588 WSQSYATTTVTPAGGDTVTIIRPNNHTVTTEWWSQSYATTTTVPAGGDTVTIIRP 645  
Db 588 SSESISSTIS-----DTNWDSSSSDLESTSIWSSS----- 620  
QY 646 IREPNNHTVTTEWWSQSYATTTTVPAGGDTVTIIRPNNHTVTTEWWSQSY 699  
Db 621 IDAQSOSQSV---SNSISTSOETSSGSEESNTSVTD-ILVSSDASSILNSD---ISS 673  
QY 700 YATTTTVPAGGDTVTIIRPNNHTVTTEWWSQSYATTTTVPAGGDTVTIIRPNN 759  
Db 674 YPSTSLSDDPHTI-AGEPDSRS-----SSSIATSTVEIS-----SDLVSLSDP- 719  
QY 760 PTVTTEWWSQSYATTTTVPAGGDTVTIIRPNNHTVTTEWWSQSYATTTTVPAGG 814  
Db 720 ----TSSFDSSSLNSDSSSP-----FSDSDIASASSFTLVAPSESLSSSS 764  
QY 815 -----RHYVNSTSDLTSESSMNTPTSISSDGMLLSTTLVTESEFTTELICSDGKE 869  
Db 765 SLSLIYPHYVNSTIYHASESESSVAP-SVASESANDOTHTL-SESDTTSIIGTDSST 822  
QY 870 ---CSRLLSSSGIVT-----NPDNESSIVTSTVPTASTMSDLSLSDGTSATSSDNV-- 919  
Db 823 VTFCRDNGOCCIVTIGTSSIDSEQTSDVTT--TSSFVASTPTSAEQSIITDNPIDS 879  
QY 920 SKSGVSVTETSVTTIOTTPNPLSSSVTSLSLQSSIPSVSESEKSVTFTNGD----NOS 975  
Db 880 SQTSSASTKSSSVSDTVVNSILLSETS--TLSSDDGTS-SDTSSISSTNSDTGGINAG 936

Qy	976	GTSDSOSTEIELVITS---STKVLPPVSSNTDLTSEPNTNR-----EOPITLSTTS	1026
Db	937	SSHKASIKESSIOKQTVTLSSSYLSLTKLSSSDITIELITTELITIEDNEPNTFTSP	996
Qy	1027	NSITEDITTSQPTGNG-----DNTSNTPNPTVATSTLASASEEDNKGSGHESATSL	1080
Db	997	SSHSIEFSS-----DNSVLKQVDRESTIKTSPTTDDVTYSSLSVHSIE-----ASTA-	1044
Qy	1081	KPSMGENS-----GLTSTIEIATTSPTAPSVSSCT-----DVTTEPDT-	1124
Db	1045	--TLGNSFNVAESTPLNATSLRSTSSSNHATE---SSGTVKSEASVEALPSPTSTD	1099
Qy	1125	-----REQPTLSTTSKTN-----SELVATTQATN-----ENGKSPSTDLSLTTGT	1168
Db	1100	NRLSYSTEEAEGITYANGSGTNLITESQVAAPTSTSVLIENLVVTSFDDNSSAAVDQ	1159
Qy	1169	SASTSANSELV---TSGSVTGGAVASANDQSHSTSV-----TNSNSI	1208
Db	1160	PSKTSIEESIMNPDSTNETNNGFIATLSQAQVPSSSHSELISTTTAKTTDASNGDSA	1219
Qy	1209	VSNTPQTLTQQVTSSTPSTNFTASTYDCSGSIHQHSTMYLGLITLLSLF	1259
Db	1220	ASNSOPTLIQVATS---SYNOPLITTYAGSSSKATKPSMLLAFISVALFF	1268
RESULT 5			
Qy	99HGK6		
AC	09HGK6	PRELIMINARY; PRT: 2297 AA.	
DT	01-MAR-2001	(Tremblrel. 16, Created)	
DT	01-MAR-2001	(Tremblrel. 16, Last sequence update)	
DE	01-DEC-2001	(Tremblrel. 19, Last annotation update)	
DE	AGGLUTININ-LIKE PROTEIN ALS7P.		
OS	ALS7.		
GN	Candida albicans (yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; Mitosporic Saccharomycetales; Candida.		
NCBI_TaxID=5476;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	SPRAIN-1161;		
RX	MEDLINE-20321177; PubMed-10861907;		
RA	Hoyer L.L., Hecht J.E.;		
RT	"The ALS6 and ALS7 genes of Candida albicans.";		
RL	Yeast 16:847-855(2000).		
DR	EMBL; AF201684; AAF98068.1; -.		
SQ	SEQUENCE 2297 AA; 244723 MW; 59B020C63027F651 CRC64;		
Query Match 39.7%; score 2576.5; DB 3; Length 2297;			
Best Local Similarity 41.3%; Pred. No. 8.7e-111;			
Matches 595; Conservative 220; Mismatches 416; Indels 211; Gaps 24;			
Qy	2	LOOPTLLFLYLSTAS--AKTITGVDFDSFNSLTWSNAANYAFKPGPYPTNVAVLGMSLDGT	59
Db	1	MKKLLYLLASFTTIVISKEVTGVFNQFNSLWSYTYRAREYESTLTANAOLEWALDGT	60
Qy	60	SANPGDTFLMPCVKFYTTTSOTSDVLTDGVKYATCFQYSGEEFTTFTLCTVNDALK	119
Db	61	IASPGDTFLMPCVKFYKMTYETSVQLTANSAYATCFDAGEDTKFSLSKCTVTDLT	120
Qy	120	SSIAFGTFTLPIAFNNGVGTSGDLSDESKCFCTAGTNTVTFNDDGKDISIDVEFEKSTVD	179
Db	121	EDTSVFGSVILPIAFNNGVGSKSTITDSCFSSGYNTVTFDGNQLSTANFLPRREL	180
Qy	180	PSAYLASRVNPSLNKVTTLFVAPCENGQYTGCTMGFSSNGDVAIDCSNIHIGIKGLN	239
Db	181	AFGLVVSQRLSMSLDMTNFVNMSTPCFMGQSGKLGFTSNDDDFEIDCSNHHVGTINEIN	240
Qy	240	DMNYPVSESESYKTKTSQNGIQIKYONPAGVRPFIDAYISATDVNOYTLAYNDVTC	299
Db	241	DMSMPVSPDPDTHICTSRALYIEFTIPAGVRPFVDIAVQIPTTEPFVFKYTFEACV	300
Qy	300	GSRLOSKEFTLRWT-CYKNSDAGSNGIVATRTVTDSTAVTTLPPNPSVDKTKIEI	358
Db	301	NGIYTSIPFSPESQPILYDALAGADLVRTTSVIGSITRTTTLPPISRLQKTKILV	360
Qy	359	LQPIPTTLITSYGVTSYLTKTAPIGATVAVDVDPYKTTTTSVSEKGTITTTTTRT	418
Db	361	LEPIPTTVTSHHGFDNIWYTKKATIGDTAIVFDVPOHTATLLTYWQESSATTTTF	420
Qy	419	NPTSDIDTVVOVPLP-----NPTVSSTE	442
Db	421	DDIDLVDTVIVKIPYENPILITTCQWNGKYLTETHEKPEPLGTDVSIKEPHNPTVTTE	480
Qy	443	YWSQSFATTTTAPPGGTDVILIREPHNHTVTTTEYWSQSFATTTTAPPGGTDVII	502
Db	481	FWSEFATTTITNYPEGTDVIVREPHNPTVTTEFWSEFATTTITNGEGTDSIV	540
Qy	503	REPPNPTVTTEYWSQSFATTTTAPPGGTDVILIREPPNPTVTTEYWSQSFATTTV	562
Db	541	REPHNPTVTTEFWSEFATTTITNGEGTDSIVREPHNPTVTTEFWSEFATTTI	600
Qy	563	TAPPGGTDVILIREPHNHTVTTTEYWSQSFATTTTAPPGGTDVILIREPHNHTVTTE	622
Db	601	TNGEGTDSIVREPHNPTVTTEFWSEFATTTITNGEGTDSIVILKEPHNPTVTTK	660
Qy	623	YWSQSFATTTTVPSPGTDVILIREPHNPTVTTEYWSQSFATTTITAPPGTDVLI	682
Db	661	FWSEFATTTITNYPEGTDVIVREPHNPTVTTEFWSEFATTTITNGEGTDSIV	720
Qy	683	REPPNHTVTTEYWSQSFATTTTAPPGGTDVILIREPHNHTVTTTEYWSQSFATTTV	742
Db	721	REPHNPTVTTEFWSEFATTTITNYPEGTDVIVREPHNPTVTTEFWSEFATTTV	780
Qy	743	TAPPGGTDVILIREPHNPTVTTEYWSQSFATTTTAPPGGTDVILIIY---ESMSSSK	798
Db	781	TNYPEGTDVIVREPHNPTVTTEFWSEFATTTITNGEGTDSIVIHDPLEESSSTA	840
Qy	799	ISTSSNDITSI-----IPS-----FSRP	816
Db	841	IESSDNISSSAQESSSVESQSFSADETSSIVELSSRSIDPSSIGLTSSESSTVSVD	900
Qy	817	HYVNSTT-----SDLSTFESSMNTPTSISSD---GMLLSSTT-----	851
Db	901	SYSSSTSESSIASYSDSYSSSIESSTLSSORYSSISDPTSPFMSDSSSDLESTITS	960
Qy	852	-----LVTESETTTELICSDGKRCRLSSSGIVTNPDNNESSIVTSTVPTAS	899
Db	961	SSIDAQSHLVQSVNSISTSQEISSSSSESTSATDALVSSDASSILSDTSYYPSS	1020
Qy	900	TMSDLSLSTDGISATSSDNVAKGVSVYTTTSTVYTTIOTTPNPLSSSVTSLTQ----	955
Db	1021	TISPDDPPTTI-AGESDSQSISITSTVEISDSVSLTSDP-ESSFDSSSLASDSSS	1078
Qy	956	PS-----VSESESKVTFTSNGDNQSGTHDSOSTEIEIVTTSSTKVLP-----	1000
Db	1079	PTDQDQDILTSSEFSFTLIKSSGSGRESSIGITLSESSDSIPTFTSTRYMSPSGMSSRHYT	1138
Qy	1001	-----VVSNT--DITSEPTNTREOPTLSTTNSITIEDITTSOPTGNDGNTSST	1049
Db	1139	NETSTSVSDVYSSVAGDETSSESVSVISESESVTSESVASESVASESVASESV	1198
Qy	1050	NPVPTVATSLASASEDNKSGHESASTSLKPSMGNSGLTTSIEATTTSTPEASP	1109
Db	1199	TAVSDI--SOLYTTSEVSTSDNSGMSP-PSSEORS-----SIPIMSSDESSESR	1249
Qy	1110	AVSSGTDVTEPTDTRQOPTLST-----TSKTNSELVATTQATNENG	1153
Db	1250	ESSSGTILSEENDS--IPTFTSTRYMSPSGMSSRHYTNTSTSTSVSDVYSSVAGDETS	1307
Qy	1154	KSPS--TDLASSLTTGTSASTANSSELVTSVSGVTCGAVASANDQSHSTSVTNSNS	1207
Db	1308	SSVSVISESESVTSESVASESVASESVTAVSDISDLVTTSEVVTSDNSNGMS	1367
Qy	1208	-----IVSNTPQ-----TTLSQQVTSSPSTNTFIATYDGCSTIOH	1245

```

Db 1368 PIPSEQRSSIPVNMSSDESSESRESSSGTILSENSDSIPT--TFSTRYLSPSGMSRRH 1425
Qy 1246 ST 1247
Db 1426 YT 1427

RESULT 6
Q9URQ0 PRELIMINARY; PRT: 468 AA.
AC Q9URQ0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN ALS2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=98440424; PubMed=9765564;
RA Hoyer L.L., Payne T.L., Hecht J.E.;
RT "Identification of Candida albicans ALS2 and ALS4 and localization of
RL als proteins to the fungal cell surface.";
RL J. Bacteriol. 180:5334-5343(1998).
DR EMBL; AF024582; AAC64237.1;
FT NON_TER 468
SQ SEQUENCE 468 AA; 50143 MW; 2E3B36D7FEAD192 CRC64;

```

```

Query Match 30.6%; Score 1986.5; DB 3; Length 468;
Best Local Similarity 80.0%; Pred. No. 2.4e-84;
Matches 375; Conservative 33; Mismatches 60; Indels 1; Gaps 1;

Qy 1 MLQQTLLFLYLSTASAKITGVDFSNLSWNAANYAFKPGYPTNNAVLGSLDGT 60
Db 1 MLQQLLLSLCVCVATKVTGVDFSNLSWNAANYAFKPGYPTNNAVLGSLDGT 60
Qy 61 ANPGDTFLNMPCKVFTTSQTSVDLTADGVKATCOFYSGEEFTFTSLCTVNDAL 120
Db 61 ANPGDTFLNMPCKVFTTSQTSVDLTADGVKATCOFYSGEEFTFTSLCTVNDAL 120
Qy 121 SIKAFGTVLPFAFNVGCTGSDLEDSKCTAGTNTVTFNDGKDISIDVEFEKSTVDP 180
Db 121 SIKAFGTVLPFAFNVGCTGSDLEDSKCTAGTNTVTFNDGKDISIDVEFEKSTVDP 180
Qy 181 SAYLYASRVMSLKNVTLFVAPQCNNGYSGTGMGSSNGDVAIDCSNIHIGITKGLND 240
Db 181 SAYLYASRVMSLKNVTLFVAPQCNNGYSGTGMGSSNGDVAIDCSNIHIGITKGLND 240
Qy 241 WNPVSSSESYTKTCTNSGIIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db 241 WNPVSSSESYTKTCTNSGIIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Qy 301 SRLOSKPFLRWGTYKNSDAGSNGIVIVATRTVTDSTTAVTLPNPSVDKTKTIELQ 360
Db 301 SRLOSKPFLRWGTYKNSDAGSNGIVIVATRTVTDSTTAVTLPNPSVDKTKTIELQ 360
Qy 361 PIPPTTTTTSVGVVTSYLTAKTAPIGETATVIVDVPYHTTTVTSEWGTITTTTTRNP 420
Db 361 PIPPTTTTTSVGVVTSYLTAKTAPIGETATVIVDVPYHTTTVTSEWGTITTTTTRNP 420
Qy 421 TDSIDTVVQVPLNPVTSTTYSQSPATTTTVPAGGCTVVIIRP 469
Db 421 TDSIDTVVQVPLNPVTSTTYSQSPATTTTVPAGGCTVVIIRP 469

```

```

RESULT 7
Q9URP8 PRELIMINARY; PRT: 468 AA.
AC Q9URP8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN ALS3.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.

```

```

ID Q9URP8 PRELIMINARY; PRT: 469 AA.
AC Q9URP8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE AGGLUTININ-LIKE PROTEIN (FRAGMENT).
GN ALS4.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=98440424; PubMed=9765564;
RA Hoyer L.L., Payne T.L., Hecht J.E.;
RT "Identification of Candida albicans ALS2 and ALS4 and localization of
RL als proteins to the fungal cell surface.";
RL J. Bacteriol. 180:5334-5343(1998).
DR EMBL; AF024586; AAC64241.1;
FT NON_TER 469
SQ SEQUENCE 469 AA; 49597 MW; 88BC96D79142C8DB CRC64;

```

```

Query Match 28.1%; Score 1823; DB 3; Length 469;
Best Local Similarity 72.1%; Pred. No. 8e-77;
Matches 338; Conservative 49; Mismatches 82; Indels 0; Gaps 0;

Qy 1 MLQQTLLFLYLSTASAKITGVDFSNLSWNAANYAFKPGYPTNNAVLGSLDGT 60
Db 1 MLQQLLLSLCVCVATKVTGVDFSNLSWNAANYAFKPGYPTNNAVLGSLDGT 60
Qy 61 ANPGDTFLNMPCKVFTTSQTSVDLTADGVKATCOFYSGEEFTFTSLCTVNDAL 120
Db 61 ANPGDTFLNMPCKVFTTSQTSVDLTADGVKATCOFYSGEEFTFTSLCTVNDAL 120
Qy 121 SIKAFGTVLPFAFNVGCTGSDLEDSKCTAGTNTVTFNDGKDISIDVEFEKSTVDP 180
Db 121 SIKAFGTVLPFAFNVGCTGSDLEDSKCTAGTNTVTFNDGKDISIDVEFEKSTVDP 180
Qy 181 SAYLYASRVMSLKNVTLFVAPQCNNGYSGTGMGSSNGDVAIDCSNIHIGITKGLND 240
Db 181 SAYLYASRVMSLKNVTLFVAPQCNNGYSGTGMGSSNGDVAIDCSNIHIGITKGLND 240
Qy 241 WNPVSSSESYTKTCTNSGIIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Db 241 WNPVSSSESYTKTCTNSGIIQKYNVPAGYRPFIDAYISATDVNQYTLAYTNDYTCAG 300
Qy 301 SRLOSKPFLRWGTYKNSDAGSNGIVIVATRTVTDSTTAVTLPNPSVDKTKTIELQ 360
Db 301 SRLOSKPFLRWGTYKNSDAGSNGIVIVATRTVTDSTTAVTLPNPSVDKTKTIELQ 360
Qy 361 PIPPTTTTTSVGVVTSYLTAKTAPIGETATVIVDVPYHTTTVTSEWGTITTTTTRNP 420
Db 361 PIPPTTTTTSVGVVTSYLTAKTAPIGETATVIVDVPYHTTTVTSEWGTITTTTTRNP 420
Qy 421 TDSIDTVVQVPLNPVTSTTYSQSPATTTTVPAGGCTVVIIRP 469
Db 421 TDSIDTVVQVPLNPVTSTTYSQSPATTTTVPAGGCTVVIIRP 469

```



QY 241 WNPVSSSESYTKCTNSNGIOIKYQNVAGYPRFDAYISATDVNQ--YTLAYTNDYTC 298  
DB 181 WNPVSSSESYTKCTNSNGIOIKYQNVAGYPRFDAYISATDVNQ--YTLAYTNDYTC 240  
QY 299 AGSLQSKPFTLWGTGKNSDAGSNGIIVATRVVDSSTAVTTLFPNPSVDKTKTIEI 358  
DB 241 TQGRKGNPLIYFWISYNSDAGSNGIIVATRVVDSSTAVTTLFPNPSVDKTKTIEV 300  
QY 359 LQPIPTTTTTSYGVVTSYLTAKTAPIGETATVIVDVP 396  
DB 301 LEPIPTTTTTSYGVVTSYLTAKTAPIGETATVIVDVP 338  
RESULT 11  
Q9HF69 PRELIMINARY; PRT: 331 AA.  
AC Q9HF69;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE AGGLUTININ-LIKE PROTEIN ALSD3P (FRAGMENT).  
OS ALSD3.  
OC Candida dubliniensis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=42374;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD36.  
RA Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Klis F.M.;  
RT "Evidence suggesting the presence of an ALS gene family in Candida  
RT dubliniensis and Candida tropicalis".  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL:AF20530; ANG3564.2; -.  
FT NON\_TER 331  
FT SEQUENCE 331 AA; 35297 MW; 469AB72F9CE0298F CRC64;

Query Match 18.1%; Score 1176; DB 3; Length 331;  
Best Local Similarity 64.3%; Pred. No. 3.4e-47;  
Matches 214; Conservative 44; Mismatches 73; Indels 2; Gaps 1;

QY 61 ANPGDTFTLNMPCVKFYTSQTSVDLTADGVKATCOFYSGEFTTSTLTCTVNDALKS 120  
DB 1 ANAGDTFTLNMPCVKFYTSQTSVDLTADGVKATCOFYSGEFTTSTLTCTVNDALKS 60  
QY 121 SIKAFGTVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 180  
DB 61 QTKALGTVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 120  
QY 181 SAYLYASRVMPKLVTLFVAPOCENGYSCTGMFSSNGDVAIDCSNHIHIGITKGLND 240  
DB 121 NDRILLSRLPSLAKSVTIFIPRCASYSSTGMFSSNGDVAIDCSNHIHIGITKGLND 180  
QY 241 WNPVSSSESYTKCTNSNGIOIKYQNVAGYPRFDAYISATDVNQYTLAYTNDYTCAG 300  
DB 181 WNPVSSSESYTKCTNSNGIOIKYQNVAGYPRFDAYISATDVNQYTLAYTNDYTCAG 238  
QY 301 SIKAFGTVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 360  
DB 239 ARPVDAGSFTNWLGVNDAAGSRIGITVTFSTVDSSTAVTTLFPNPSVDKTKTIEIQL 298  
QY 361 PIPTTTTTSYGVVTSYLTAKTAPIGETATVIV 393  
DB 299 PIPTTTTTSYGVVTSYLTAKTAPIGETATVIV 331

RESULT 12  
Q9Y8F1 PRELIMINARY; PRT: 433 AA.  
AC Q9Y8F1;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE AGGLUTININ-LIKE PROTEIN 7 (FRAGMENT).  
GN ALS7.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RA MEDLINE-20321177; PubMed=10861907;  
RA Hoyer L.L., Hecht J.E.;  
RT "The ALS6 and ALS7 genes of Candida albicans.";  
RT Yeast 16:847-855(2000).  
DR EMBL:AF075294; AAD42034.1; -.  
FT NON\_TER 433  
FT SEQUENCE 433 AA; 48037 MW; 1F30CFD99C2EC445 CRC64;

Query Match 15.4%; Score 999.5; DB 3; Length 433;  
Best Local Similarity 46.0%; Pred. No. 6e-39;  
Matches 199; Conservative 76; Mismatches 155; Indels 3; Gaps 2;

QY 2 LQOFTLLFLYLSTAS--AKTITGVDFSNLSLWNAANYAFKPGCYPTWNAVLGWSLDGT 59  
DB 1 MKKLYLLYLLASEFTVISKEVTGVNQNLSIYRYRYEISTLTANAQLEWALDGT 60  
QY 60 SANPCDFTLNMPCVKFYTSQTSVDLTADGVKATCOFYSGEFTTSTLTCTVNDALK 119  
DB 61 IASPDFTLNMPCVKFYTSQTSVDLTADGVKATCOFYSGEFTTSTLTCTVNDALK 120  
QY 120 SIKAFGTVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 179  
DB 121 ETSVFGSVTLPIAFNVGGTSGSTDLSDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDP 180  
QY 180 PSAYLYASRVMPKLVTLFVAPOCENGYSCTGMFSSNGDVAIDCSNHIHIGITKGLN 239  
DB 181 AGLVVSORLSMLDITNFVMSPTCFMGYSGLGFTSNDGDDFDCSSIHVGTNEIN 240  
QY 240 WNPVSSSESYTKCTNSNGIOIKYQNVAGYPRFDAYISATDVNQYTLAYTNDYTCAG 299  
DB 241 DMSMVSSVPFDHTICTSRALYIEFTIPAGYRPFVDAIVQIPTTEFFVKYTNFACV 300  
QY 300 GSRLOSKPFTLRWT-GYKNSDAGSNGIIVATRVVDSSTAVTTLFPNPSVDKTKTIEI 358  
DB 301 NGIYTSIPFTSFSQPILYDEALAGADLVRTTSTVIGSITRTTLPFISRLQTKTILV 360  
QY 359 LQPIPTTTTTSYGVVTSYLTAKTAPIGETATVIVDVPYVHTTSTVSEMTGTTITTTT 418  
DB 361 LEPIPTTTTTSYGVVTSYLTAKTAPIGETATVIVDVPYVHTTSTVSEMTGTTITTTT 420  
QY 419 NPTDSIDTVVQV 431  
DB 421 DDIDLVDIVIVKI 433

RESULT 13  
Q9HF71 PRELIMINARY; PRT: 353 AA.  
AC Q9HF71;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE AGGLUTININ-LIKE PROTEIN 1 (FRAGMENT).  
GN ALST1.  
OS Candida tropicalis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5482;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13803;





09HG16  
 ID Q9HG16 PRELIMINARY; PRT; 226 AA.  
 AC Q9HG16  
 DT 01-MAR-2001 (TREMELrel. 16, Created)  
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
 DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)  
 DE AGGLUTININ-LIKE PROTEIN (FRAGMENT).  
 GN ALS.  
 OS Candida albicans (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5476;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1177;  
 RA Hoyer L.L., Kapteyn J.C., Hecht J.E., En J., Klis F.M.;  
 RT "Evidence Suggesting the Presence of an ALS Gene Family in Candida  
 RT dubliniensis and Candida tropicalis";  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF189016; AAG17111.1;  
 DR NON\_TER 1  
 DR NON\_TER 226  
 SQ SEQUENCE 226 AA; 24802 MW; 2AA36A42E44CB460 CRC64;

Query Match 12.78; Score 824; DB 3; Length 226;  
 Best Local Similarity 70.18; Pred. No. 3.7e-31;  
 Matches 155; Conservative 28; Mismatches 38; Indels 0; Gaps 0;

QY 429 VQVPLPNTVSTTEYWSQSFATTTTAPPGTDTVIIRPPNHTVTTEYWSQSFATTT 488  
 DB 1 VQVSPNPTVTTFQWSSGVPTTETVTGQGTDSVIRKEPHNPTVTTFWSEFATTE 60  
 QY 489 TVTAPPGTDSVIRPPNPTVTTEYWSQSFATTTTAPPGTDSVIRPPNPTVT 548  
 DB 61 TVTNPEGTDSVIRKEPHNPTVTTFWSEFATTTETVTNYPEGTDVIRPPNPTVT 120  
 QY 549 TEYWSQYATTTTAPPGTDSVIRPPNHTVTTEYWSQYATTTTAPPGTDTV 608  
 DB 121 TEFWSEFATTTETVTNYPEGTDVIRPPNPTVTTFWSEFATTTETITNYPEGTDV 180  
 QY 609 IREPPNHTVTTEYWSQSFATTTTVPSTGTDVIRPP 649  
 DB 181 IVREPHNPTVTTFWSESFVTTTITGTLGTDIVINDP 221

Search completed: October 3, 2002, 15:37:25  
 Sp time: 306 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:25:37 ; Search time 31.64 Seconds  
(without alignments)  
1460.389 Million cell updates/sec

Title: US-09-715-876-8\_COPY\_17\_432  
Perfect score: 2190  
Sequence: 1 AKTITGVDFNSLWSNA.....TTTTTRNPTSDITVVQVP 416

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	308.5	14.1	650	15	AA197575
2	177	8.1	1537	15	AA197575
3	175.5	8.0	1837	21	AA111726
4	175	8.0	1721	19	AA148299
5	175	8.0	1721	21	AA111727
6	156	7.1	688	22	AB130137
7	156	7.1	688	22	AB135307
8	156	7.1	688	22	AB120749
9	156	7.1	688	22	AA156138
10	156	7.1	688	22	AA168511
11	156	7.1	688	22	AA116315

12	156	7.1	688	22	AA128810	Peptide #2847 enco
13	156	7.1	688	22	AA104053	Peptide #2735 enco
14	155	7.1	894	15	AA197578	Flocculation prote
15	155	7.1	894	15	AA197578	S. cerevisiae FLO1
16	153.5	7.0	1795	22	AB169806	Drosophila melanog
17	151	6.9	2586	22	AB169806	C. parvum Iowa iso
18	150.5	6.9	175	21	AA111743	T. gondii immunoge
19	150.5	6.9	288	20	AA129081	T. gondii immunoge
20	150.5	6.9	288	20	AA129082	T. gondii immunoge
21	150.5	6.9	288	20	AA129082	T. gondii immunoge
22	150.5	6.9	288	20	AA129082	T. gondii immunoge
23	150	6.8	386	22	AB129832	Peptide #2483 enco
24	150	6.8	386	22	AB129832	Peptide #2514 enco
25	150	6.8	386	22	AB129832	Protein #2421 enco
26	150	6.8	386	22	AA15821	Human bone marrow
27	150	6.8	386	22	AA15821	Peptide #2449 enco
28	150	6.8	386	22	AA15821	Peptide #2553 enco
29	150	6.8	386	22	AA15821	Peptide #2431 enco
30	150	6.8	386	22	AA15821	Cryptosporidium pa
31	148.5	6.8	216	21	AA111729	Yeast 2.6 KB agglu
32	148.5	6.8	862	15	AA10563	C. parvum NINC iso
33	146.5	6.7	243	21	AA111746	Novel human diagno
34	146.5	6.7	1045	22	AB108332	Human ORFX ORF709
35	146.5	6.7	1532	21	AA10945	C. parvum Iowa iso
36	145	6.6	130	21	AA111740	Mutant protease (d
37	145	6.6	1959	12	AA10562	Mutant protease (K
38	145	6.6	1962	12	AA10560	Mutant protease (N
39	145	6.6	1962	12	AA10561	Mutant protease (A
40	145	6.6	1962	12	AA10557	Mutant protease (A
41	145	6.6	1962	12	AA10558	Mutant protease (A
42	145	6.6	1962	12	AA10559	Mutant protease (K
43	145	6.6	1962	12	AA10563	Mutant protease (d
44	145	6.6	1962	12	AA10941	Mutant protease (d
45	145	6.6	1974	12	AA10940	Mutant protease (d

## ALIGNMENTS

RESULT 1  
AA197575  
ID AA197575 standard; Protein; 650 AA.  
XX  
AC AA197575;  
DT 19-JUL-1994 (first entry)  
XX  
DE Alpha-agglutinin of Saccharomyces cerevisiae.  
XX  
KW Immobilisation; enzyme; cell wall; alpha agglutinin; AGA 1; FLO 1;  
KW Major cell wall protein; glycosyl-phosphatidyl-inositol;  
KW anchoring protein; alpha factor; alpha-agglutinin; invertase;  
KW inulinase; alpha-amylase; Saccharomyces cerevisiae;  
KW enzymatic process; fermentation; biodegradation; catalysis.  
XX  
OS Saccharomyces cerevisiae.  
XX  
PN WO9401567-A.  
XX  
PD 20-JAN-1994.  
XX  
PF 07-JUL-1993; 93WO-EP01763.  
PR 08-JUL-1992; 92EP-0202080.  
PR 14-DEC-1992; 92EP-0203899.  
PA (UNIL ) UNILEVER NV.  
PA (UNIL ) UNILEVER NV.  
PI Kils FM, Schreuder MP, Toschka H, Verrips CT;  
XX WPI; 1994-035071/04.  
DR N-PSOB; AAQ54012.

XX Immobilisation of enzymes to microbial cell wall - by prodn. of  
PT fusion protein of enzyme linked to anchoring protein  
XX  
XX  
PS Example 1; Page 32-39; 99pp; English.  
XX  
XX The alpha-agglutinin is used in a method to immobilise enzymes to a  
CC microbial cell wall. The coding sequence is used in the production  
CC of a recombinant polynucleotide which comprises a structural gene  
CC encoding a protein with catalytic activity and at least part of a  
CC gene encoding at least the C-terminus of a protein capable of  
CC anchoring in a eukaryotic or prokaryotic cell wall. The anchoring  
CC fragment or protein is selected from alpha agglutinin, AGA 1, FLO 1,  
CC major cell wall protein of lower eukaryotes or a proteinase of  
CC lactic acid bacteria. The recombinant polynucleotide preferably  
CC also comprises a sequence encoding a signal peptide to ensure  
CC secretion of the expressed product. The signal peptide is  
CC preferably derived from glycosyl-phosphatidyl-inositol anchoring  
CC protein, alpha factor, alpha agglutinin, invertase or inulinase,  
CC alpha-amylase of bacillus or proteinases of lactic acid bacteria.  
CC The host microorganism can be used for performing enzymatic  
CC processes on an industrial scale.  
XX  
XX Sequence 650 AA;  
SQ

Query Match 14.1%; Score 308.5; DB 15; Length 650;  
Best Local Similarity 25.8%; Pred. No. 2.7e-16;  
Matches 116; Conservative 79; Mismatches 190; Indels 65; Gaps 20;

- QY 10 SPNSLTWSN-----AANVAFKPGCGYPTWNAVGLWSL-DGTSANPGDFTFLNMPGVFK-- 60  
Db 21 nldntfslneleplcan---kqpd-ggwcatdfsladassiregdeflsmphvyrlik 76  
QY 61 -VYTSQTSVDLTADGVKATQFVSGE-----EFTTFSTLCTVNDALKSSIKAFQTVTL 114  
Db 77 llnsqqtalsladgfeafkc-yvsqqaelyenttf---tctaquadlssyntidgsitf 132  
QY 115 PIAFNVGGTSGSDLEDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDPASV---LYAS 171  
Db 133 slntsdggsyeyelenakfksgmlvklngmsdv---vnf-----dpaatenvfhs 184  
QY 172 RVMSLNKRVTLFVAPQCENGYTSG---TMGFSSSNGDVAIDCSNIHIGITKGLNDMNPV 228  
Db 185 grstgysgfesylhgmcpngyflggtetekldydsnnnnvldcsavqvsydsndfndwfp 244  
QY 229 VSSEFSYTKTCTSGNIQIKY-QNVAPGYRPFIDAYIS-ATDVMQYTLAYTNDYTC--- 282  
Db 245 qsyndtnadvctcfnlwtideklydgemlwnalqslpanvntidhaldefgtycltdi 304  
QY 283 AGSRLQSKPFTLR---WTGKNSDAGSNGIVIVATRTVTD-----STVATVTLTFNP 332  
Db 305 anttyatqgfettrefivvggrnlgtsakssfsttttdltsintaysygsistv--- 360  
QY 333 SVDKTKTIEILOPIPTTTTTSYGVVTSYLTKTAPIGETATVIVDVPYHTTT----- 385  
Db 361 etgnrttceevishvvtstkslptatsltiaqtslystdsnltvgdihntsevisdve 420  
QY 386 TVTSEWGTGTTTT-TTRTNPTDSIDTVVVO 414  
Db 421 tistretastvvaaptsttgwgamntyipq 450

RESULT 2  
ID AAR60562 standard; Protein; 1537 AA.  
AC AAR60562;  
XX

DT 13-APR-1995 (first entry)

XX Yeast 4.7 kb agglutination gene FLO1L.  
DE  
XX

KW Yeast; agglutination; FLO1L.  
XX Saccharomyces cervisiae ABXL-ID.  
OS  
XX WO9419475-A.  
PN  
XX 01-SEP-1994.  
PD  
XX 24-FEB-1994; 94WO-JP00290.  
PF  
XX 26-FEB-1993; 93JP-0038871.  
PR  
XX (PANI-) PANIMOLABORATORIO BRYGERILABORATORIUM.  
PA (SAPB) SAPPORO BREWERIES.  
XX Kersenen S, Ogawa M, Onnela M, Penttila M, Takata Y;  
PI Watari J;  
XX WPI: 1994-294338/36.  
DR N-PSDB; AAQ71390.  
XX New yeast agglutination genes and yeast contg. them - impart  
PT agglutination properties to facilitate removal from fermentation  
PT media  
XX  
XX Disclosure; Page 43-48; 75pp; English.  
XX The agglutination gene is called FLO1. Saccharomyces cerevisiae  
CC includes an agglutination gene of 4.7 kb (FLO1L) and an  
CC agglutination gene of 2.6 kb (FLOIS). FLO1L is the intact FLO1 gene  
CC on chromosome I, and FLOIS is the FLO1L gene with a portion of the  
CC ORF deleted in frame. FLO1L imparts a relatively strong  
CC agglutinative property to the host yeast into which it is introduced,  
CC while FLOIS imparts a weaker agglutinative property.  
XX  
XX Sequence 1537 AA;  
SQ

Query Match 8.1%; Score 177; DB 15; Length 1537;  
Best Local Similarity 23.0%; Pred. No. 3.5e-05;  
Matches 112; Conservative 47; Mismatches 181; Indels 146; Gaps 25;

QY 21 NYAFKPGC-YPTWNAVGLWSL--GTSANPGDFTFLNM-----PCVFKYTSQ-- 65  
Db 104 nwgckgmgaacsnggiaywstldfgyttptnv-tiemtyflppqtgsytkfatvdds 162  
QY 66 -----TSVDLTADGVK-----YATCFYSGEEFTTFTSL 94  
Db 163 ailsvggatafnccagqgppitstnftldgikpwggslppniegtvymegyyvpm---- 218  
QY 95 TCTVNDALKSSIKAFGT---VTLPIAFNVGGTSGSDLEDSKCFAGTNTVTFNDGDKD 150  
Db 219 -----kvvysnavswgtlplsvtlp-----dgttvsddfg-----yyvsfdd--- 256  
QY 151 ISIDVEFEKSTV-DPSAYLYASRVMPSLNKVTTLFVAPQCENGYTSGMTGFSSSNGDVAI 209  
Db 257 ---disqncetvpdpnsya-vsttttttpewtgtgtststemttvtgtngvptdetvivi 312  
QY 210 DCSNIHIGITKGLNDMNPVSSSEFSYTKTCTSGNIQIKYQNVAPGYRPFIDAYISATDV 269  
Db 313 rtpetastilttepewnstftsteltvtgtngvrtdei-----virtpttatta 366  
QY 270 NQYTLAYTNDYTCAGSLQSKPFTLRWTKYKNSDAGSNGI---VIVATRTVTDSTAVT 326  
Db 367 itttepwnstftststel-----ttvtgtnglptdetilivrtptattant 413  
QY 327 TL-PENPSVDKTKT-----IEILOPIPTTTT-----SVGVVTS 361  
Db 414 tctqpwndtftststemttvtgtnglptdetilivrtptattantttgwdtftstste 473  
QY 362 YLTTKTA---PIGETATVIVDVPYHTTT--TVTSEWGTGTT-----TTTTTRN--PTDS 407  
Db 474 mttvtgtnglptdet-llivrtptattantttqpwndtftststemttvtgtnglptd- 531





Db 632 --kpiatttttkpivtttttkattttttvptttt-ttkrdemttttplp 681

RESULT 6

ID ABB30137 standard; Peptide: 688 AA.

XX ABB30137;

DT 01-FEB-2002 (first entry)

XX Peptide #2788 encoded by breast cell single exon nucleic acid probe.

DE Human; microarray: single exon probe; gene expression; breast;

KW disease: cancer.

XX Homo sapiens.

XX WO200157271-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00662.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234497.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-496933/54.

DR New spatially-addressable set of single exon nucleic acid probes,

PT useful for measuring gene expression in sample derived from human

PT breast, comprises number of single exon nucleic acid probes -

XX Claim 27; SEQ ID NO 13105; 327pp + sequence listing; English.

XX The invention relates to a spatially-addressable set of single exon

CC nucleic acid probes for measuring gene expression in a sample derived

CC from human breast and BT 474 cells. The method involves contacting

CC the probes with a collection of detectably labelled nucleic acids

CC derived from mRNA of human breast, and then measuring the label

CC bound to each probe of the microarray. The probes are useful for

CC verifying the expression of regions of genomic DNA predicted to

CC encode proteins. They are useful for gene discovery, and for

CC determining predisposition and/or prognosing breast disease. Gene

CC expression analysis is useful for assessing the toxicity of chemical

CC agents on cells. The microarray of this invention presents a far greater

CC diversity of probes for measuring gene expression, with far less bias

CC than expressed sequence tag microarrays. The method is suitable for

CC rapid production of functional information from genomic sequence. The

CC present sequence is a peptide encoded by a single exon nucleic acid

CC probe of the invention.

CC Note: The sequence data for this patent did not form part of the

CC printed specification, but was obtained in electronic format directly

CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).

XX Sequence 688 AA;

XX

Query Watch 7.18; Score 156; DB 22; Length 688;

Best Local Similarity 22.74; Pred. No. 0.00055;

Matches 93; Conservative 64; Mismatches 176; Indels 76; Gaps 16;

QY 43 TSANPGDTFTLMPNCVFKYTTTSQTSVDLTADG-----VKYATCFYSGEFTTTF 91

Db 6 tslegsettsva-----tgsettvstegsgttvtstgsettkvat-----tgsetttt 55

QY 92 STLTCTVNDALKSIKAFGVTLPIAFNVGGTSGSTDLSDSKCFAGTACTNTVTFNDGDXDI 151

Db 56 stegselitta-----aitgsetttastegsetttastegsettsastgsetttasttsset 112

QY 152 SI-DVEFEKSTVDPVSAYLYASRVMPSLNKKVTTLFVAPQCENGYSGTGMGFFSSNGDVAID 210

Db 113 tmasimgsettmastgsettkvatssakmttvtft-----ensttasttastsetts-- 166

QY 211 CSNIHIGITKGLNDWNYPVSSEFSYTKTCTSNIGIQIKYQNPAGYRPFIDAYISATDVN 270

Db 167 -----tagse--tipastagsetttttstegsetttastegs-----ettastess 211

QY 271 QYTLAVT-NDYTCAGSRLQSKPFTLRWTKYKNSDAGSNGIVIVATRTVTDSTTA----- 324

Db 212 ettattigsetttastegsetttstegsetttastegsetttstgsetttastegs 271

QY 325 -----VTTLPFNPSVDKTKTIER-LQPIPTTTTITSYGVVTT-----SYLTAKTA 367

Db 272 ettastegselttv-----stgsetltvasagsetttvtimgsetttastagsetts 327

QY 368 PIG-ETATVIVDVPYHHTTIVTSEWTCITTTTTRT--NPTDSDTIVVV 413

Db 328 tagsetttastegsettsstgsetttvtgtcttstegsetttv 376

RESULT 7

ABB35307

ID ABB35307 standard; Peptide: 688 AA.

XX ABB35307;

XX 04-FEB-2002 (first entry)

DE Peptide #2813 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234497.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

PI WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human foetal liver -

XX Claim 27; SEQ ID NO 27942; 639pp + sequence listing; English.

XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal

CC liver. The single exon nucleic acid probes may be used for predicting,

CC measuring and displaying gene expression in samples derived from human

CC foetal liver. The present sequence is a peptide encoded by a single exon

CC nucleic acid probe of the invention.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 688 AA;

Query Match 7.1%; Score 156; DB 22; Length 688;  
Best Local Similarity 22.7%; Pred. No. 0.00055;  
Matches 93; Conservative 64; Mismatches 176; Indels 76; Gaps 16;  
QY 43 TSANPGDTFTLNPCKVFKYTSQTSVLDLTADG-----VKYATCFYSGEFTTF 91  
Db 6 tategsetttvsa-----tgsetttvstgsgttvtlsgsettkvst----tgsetttt 55  
QY 92 STLTCTVNDALKSSIKAFGTVTLPFAFNVGCTGSSDLEDSKCFCTAGTNTVTFNDGDKDI 151  
Db 56 stegseitta---slgsetttastegsetttastegsetttastegsetttastet 112  
152 SI-DVEFEKSTVPDSAYLVASRVMPSLNKVTLFVAPQCENGYTSCTMGFSSNGDVAID 210  
Db 113 tmasimgsettmastigsettkvtasakmtvft----ensttiastasetttvs-- 166  
QY 211 CSNIHIGITKGLNDWNPVSSFSYTKTCTNSGQIKYONVPAGYRPFIDAVISATDVN 270  
Db 167 -----tagse--tipastagsettttstegsetttastegs-----ettastess 211  
QY 271 QYTLAYT-NDYTCAGSRLOSKPFLRTWTGYNKSDAGSNGIVIVATRTVTDSTTA----- 324  
Db 212 ettattigsetttastegsetttstegsetttastegsetttstgsetttastegs 271  
QY 325 -----VTLFPNPSVDKTKTIEI-LQPIPTTTTTSYGVVTT-----SYLTAKTA 367  
Db 272 ettastegsetttv-----stgsetttvsaegsetttvtmgsetttastagsetttvs 327  
QY 368 FIG-ETATVIVDPYHTTTVTSEWTCITTTTTRT--NPTDSIDTVV 413  
Db 328 tagsetttastegsetttvstgsetttvtgttettstegsetttv 376

RESULT 8  
ID ABB20749  
XX ABB20749 standard; Protein; 688 AA.  
AC ABB20749;  
XX  
DT 23-JAN-2002 (first entry)  
Protein #2748 encoded by probe for measuring heart cell gene expression.  
KW Human: gene expression; heart; microarray; vascular system;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease.  
XX  
XX Homo sapiens.  
XX  
XX WO200157274-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00666.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.  
XX  
XX 21-SEP-2000; 2000US-0234687.  
XX  
XX 27-SEP-2000; 2000US-0236359.  
XX  
XX 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI

XX  
DR WPI; 2001-488899/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
PS Claim 15; SEQ ID NO 22519; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart (see  
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such  
CC probe. The probes may be used for predicting, measuring and displaying  
CC gene expression in samples derived from the human heart via microarrays.  
CC By measuring gene expression, the probes are useful for predicting,  
CC diagnosing, grading, staging, monitoring and prognosing diseases of the  
CC human heart and vascular system e.g. cardiovascular disease,  
CC hypertension, cardiac arrhythmias and congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 688 AA;

Query Match 7.1%; Score 156; DB 22; Length 688;  
Best Local Similarity 22.7%; Pred. No. 0.00055;  
Matches 93; Conservative 64; Mismatches 176; Indels 76; Gaps 16;  
QY 43 TSANPGDTFTLNPCKVFKYTSQTSVLDLTADG-----VKYATCFYSGEFTTF 91  
Db 6 tategsetttvsa-----tgsetttvstgsgttvtlsgsettkvst----tgsetttt 55  
QY 92 STLTCTVNDALKSSIKAFGTVTLPFAFNVGCTGSSDLEDSKCFCTAGTNTVTFNDGDKDI 151  
Db 56 stegseitta---slgsetttastegsetttastegsetttastegsetttastet 112  
QY 152 SI-DVEFEKSTVPDSAYLVASRVMPSLNKVTLFVAPQCENGYTSCTMGFSSNGDVAID 210  
Db 113 tmasimgsettmastigsettkvtasakmtvft----ensttiastasetttvs-- 166  
QY 211 CSNIHIGITKGLNDWNPVSSFSYTKTCTNSGQIKYONVPAGYRPFIDAVISATDVN 270  
Db 167 -----tagse--tipastagsettttstegsetttastegs-----ettastess 211  
QY 271 QYTLAYT-NDYTCAGSRLOSKPFLRTWTGYNKSDAGSNGIVIVATRTVTDSTTA----- 324  
Db 212 ettattigsetttastegsetttstegsetttastegsetttstgsetttastegs 271  
QY 325 -----VTLFPNPSVDKTKTIEI-LQPIPTTTTTSYGVVTT-----SYLTAKTA 367  
Db 272 ettastegsetttv-----stgsetttvsaegsetttvtmgsetttastagsetttvs 327  
QY 368 FIG-ETATVIVDPYHTTTVTSEWTCITTTTTRT--NPTDSIDTVV 413  
Db 328 tagsetttastegsetttvstgsetttvtgttettstegsetttv 376

RESULT 9  
ID AAM56138  
XX AAM56138 standard; Protein; 688 AA.  
AC AAM56138;  
XX  
DT 05-NOV-2001 (first entry)  
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 28243.  
XX Human: brain expressed exon; gene expression analysis; probe;  
XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
XX epilepsy; cancer.  
XX  
XX Homo sapiens.  
XX





```
OY 368 PIG-ETATVIVDPYHTTTTSEWGTGTTTTRT--NPTDSIDTVV 413
||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 328 tagsettaiegsetttvstgsetttvtgttettstgsetttv 376

RESULT 11
AAM16315
ID AAM16315 standard; Protein; 688 AA.
XX
AC AAM16315;
XX
DT 12-OCT-2001 (first entry)
DE Peptide #2749 encoded by probe for measuring cervical gene-expression.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW Cervical cancer.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 27; SEQ ID No 21141; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded
CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC can be used to produce a single exon microarray, which can be used for
CC measuring human gene expression in a sample derived from human cervical
CC epithelial cells. By measuring gene expression the probes are therefore
CC useful in grading and/or staging of diseases of the cervix, notably
CC cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 688 AA:

Query Match 7.1%; Score 156; DB 22; Length 688;
Best Local Similarity 22.7%; Pred. No. 0.00055;
Matches 93; Conservative 64; Mismatches 176; Gaps 16;

OY 43 TSANPGDTFLNMPVCVKYTTTSQTSVDLTADG-----VKYATCQFYSGEFTTF 91
|| : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 tategsetttvsa-----tgsetttvsgsgttvtstgsettkvst---tgsetttt 55

OY 92 SFLCTVNDALKSKIKAFCTVTLFIATNVGCTGSDTDESKCFCTAGTNTVTFNDGDKDI 151
|| : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 stegseitta---siltgsetttastgsetttastgsetttasttsst 112
```





Qy 259 FIDAVISATDVNQYTLAYTNDYTCAGSRLOSKPFTLRWTGKNSDAGSNGIGIVIVATRTV 318  
Db 316 tseglstt-----tepwgtgfts-tstevttltgtnqgp 349  
Qy 319 TDSTTAVTTLPPNPSVDKTKTIELQPIPTTTTTSYGVVTSYLTKTAPIGETATVI-V 377  
Db 350 tdetvirltptseglsttt---epw-tgftststemttvtgtnqgptdetvirlt 404  
Qy 378 DVPYHTTTTTSWTGCTTTTTR-----TN--PTDSIDTVVVQVP 416  
Db 405 ptseglvttttepwtgtftststemstvtgtnlptde-tviviuktp 450

Search completed: October 3, 2002, 16:29:43  
Job time: 246 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:27:32 ; Search time 14.64 Seconds  
(without alignments)  
694.061 Million cell updates/sec

Title: US-09-715-876-8\_COPY\_17\_432

Perfect score: 2190  
Sequence: 1 AKITGVDFSNLSLTSNAA.....TTTTRTNPTDSDTIVVQVP 416

Scoring table: BLOSUM62

Gepop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCUS-COMB.pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	308.5	14.1	650	3	US-08-362-525-2
2	177	8.1	1537	1	US-08-325-267A-2
3	175.5	8.0	1837	3	US-08-928-361B-5
4	175	8.0	1721	3	US-08-700-651-5
5	175	8.0	1721	3	US-08-928-361B-6
6	155	7.1	894	3	US-08-362-525-22
7	155	7.1	894	3	US-08-971-692-15
8	150.5	6.9	175	3	US-08-700-651-12
9	150.5	6.9	175	3	US-08-928-361B-17
10	148.5	6.8	216	3	US-08-928-361B-8
11	148.5	6.8	862	1	US-08-325-267A-4
12	146.5	6.7	249	3	US-08-700-651-15
13	146.5	6.7	249	3	US-08-928-361B-20
14	145	6.6	130	3	US-08-700-651-9
15	145	6.6	130	3	US-08-928-361B-14
16	143.5	6.6	216	3	US-08-928-361B-27
17	140	6.4	130	3	US-08-700-651-8
18	140	6.4	130	3	US-08-928-361B-13
19	140	6.4	138	3	US-08-700-651-10
20	140	6.4	138	3	US-08-928-361B-15
21	139.5	6.4	150	3	US-08-928-361B-18
22	139	6.3	128	3	US-08-700-651-7
23	139	6.3	128	3	US-08-928-361B-12
24	138.5	6.3	162	3	US-08-700-651-13
25	136	6.2	1477	1	US-08-038-582-4
26	136	6.2	1477	1	US-08-303-532-4
27	136	6.2	1477	2	US-08-530-198-4

28	136	6.2	1477	2	US-08-469-880-4	Sequence 4, Appl
29	136	6.2	1477	2	US-08-728-470-4	Sequence 4, Appl
30	136	6.2	1477	2	US-08-617-697-4	Sequence 4, Appl
31	136	6.2	1477	2	US-08-719-641-4	Sequence 4, Appl
32	134.5	6.1	124	3	US-08-700-651-11	Sequence 11, Appl
33	134.5	6.1	124	3	US-08-928-361B-16	Sequence 16, Appl
34	134	6.1	1222	2	US-08-682-517-15	Sequence 15, Appl
35	134	6.1	1222	2	US-08-682-517-9	Sequence 9, Appl
36	132	6.0	629	5	US-09-241-581B-6	Sequence 6, Appl
37	132	6.0	629	5	PCT-US95-07721-6	Sequence 6, Appl
38	131.5	6.0	562	6	5258502-2	Patent No. 5258502
39	128	5.8	907	3	US-08-783-774-2	Sequence 2, Appl
40	128	5.8	907	5	PCT-US95-04611A-19	Sequence 19, Appl
41	127.5	5.8	1426	4	US-09-136-574A-43	Sequence 43, Appl
42	126.5	5.8	451	1	US-08-287-001A-2	Sequence 2, Appl
43	126.5	5.8	451	5	PCT-US95-09941-2	Sequence 2, Appl
44	120	5.5	806	1	US-08-270-076A-11	Sequence 11, Appl
45	119.5	5.5	2314	4	US-09-268-347-49	Sequence 49, Appl

## ALIGNMENTS

RESULT 1  
US-08-362-525-2  
; Sequence 2, Application US/08362525  
; Patent No. 6027910  
; GENERAL INFORMATION:  
; APPLICANT: ALIIS, FRANCISCUS M.  
; APPLICANT: SCHREUDER, MARTEN P.  
; APPLICANT: TOSCHKA, HUIJSE Y.  
; APPLICANT: VERRIPS, CORNELIS T.  
; TITLE OF INVENTION: PROCESS FOR IMMOBILIZING ENZYMES TO THE  
; TITLE OF INVENTION: CELL WALL OF A MICROBIAL CELL BY PRODUCING A FUSION  
; TITLE OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DABY & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D. C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/08/362,525  
; FILING DATE: 04-JAN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92202080.5  
; FILING DATE: 08-JUL-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: EP 92203899.7  
; FILING DATE: 14-DEC-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/EP93/01763  
; FILING DATE: 07-JUL-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KOKULIS, PAUL N.  
; REGISTRATION NUMBER: 16,773  
; REFERENCE/DOCKET NUMBER: 213289/T7020(V)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 822-0944  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 650 amino acids  
; TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-362-525-2

Query Match 14.1%; Score 308.5; DB 3; Length 650;  
Best Local Similarity 25.8%; Pred. No. 1.8e-18;  
Matches 116; Conservative 79; Mismatches 190; Indels 65; Gaps 20;

QY 10 SPNSLTWSN-----AANYAFKPGCYPTWNAVGLWSL-DGTSANPGDFTLNPCKVFK-- 60  
DB 21 NINDITESLEITPLTAN---KQPD-QGWTATDFSIADASSIREGDEFTLSMPHYRIK 76  
QY 61 YTFSTQSVOLTAGVKYATCOFYSGE-----BFTTFSLCTVNDALKSSIKAFCTVTL 114  
DB 77 LNSSQTATSLADGTEAFKC-YYSQQAALYENTTF---TCTAQNDLSSYNTIDSGISIF 132  
QY 115 PIAFNVGCTGSSDLEDSKCTAGTNTVTFNDGDKDISIDVEFEKSVDPISAY---LYAS 171  
DB 133 SLNFSQDSSSYEYLENAKFKSGMLVKLGNQMSDV---VNF-----DPAFTENVFHS 184  
QY 172 RVMSFLNKVTVLVAPOCENGYSG---TWGFSSSNGDVAIDCSNIHIGITKGLNDWNP 228  
DB 185 GRSTGSGSFESYHLGMYCPNGFLGTEKIDIDSSNNVDDLCSSVQVYSSNDFNDWFP 244  
QY 229 VSSSGSFYKTCNSGKIYK-ONVPAGYRPFIDAVIS-ATDYNQYFLATNDYTC---- 282  
DB 245 QSYNDTNADVTCGSLWLTIDKLYIGEMLVNALQSLPANVNTIDHLEFOYTCCLDTI 304  
QY 283 AGSLRQSKPTLR---WTGYKNSDAGSNGIVIVATRTVTD-----STTAVTTLPPNP 332  
DB 305 AMTYATQFSTTEFTVYOGRLGTASAKSFISTTTTDLTSINTSAYSGSISTV---- 360  
QY 333 SVDKTKTEILOPIPTTTTTSYGVGVTTSYLTAPIGETATVVDVVPYHHTT----- 385  
DB 361 EFGNRTTSEVISHVVTTSKPSPTATSLIAQTSYTSNITVGTDIHTTSEIVSDVE 420  
QY 386 VYTSWGTGTTT-TTRNPNPDSIDTVVQV 414  
DB 421 TISRETASTVAAPTSTTGWTGAMTVIPQ 450

RESULT 2  
US-08-325-267A-2  
Sequence 2, Application US/08325267A  
Patent No. 5583271

GENERAL INFORMATION:  
APPLICANT: WATARI, JUNJI  
APPLICANT: TAKATA, YOSHIHIRO  
APPLICANT: OGAWA, MASAHIRO  
APPLICANT: PENTILA, MERJA  
APPLICANT: ONNELA, MAIJA-LEENA  
APPLICANT: KERANEN, SIRKKA  
TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST  
TITLE OF INVENTION: CONTAINING THEM  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400  
CITY: ARLINGTON  
STATE: VA  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/325,267A  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP PCT/JP94/00290  
FILING DATE: 24-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 38871/1993  
FILING DATE: 26-FEB-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2589-023-0XPCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1537 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-325-267A-2

Query Match 8.1%; Score 177; DB 1; Length 1537;  
Best Local Similarity 23.0%; Pred. No. 1.5e-06;  
Matches 112; Conservative 47; Mismatches 181; Indels 146; Gaps 25;

QY 21 NTAFKPGC-YPTWNAVGLWSL-DGTSANPGDFTLNM-----PCVFKYTSQ-- 65  
DB 104 NMCKGMGACNSQGIAYWSTDLFGYPTTNV-TLEMGYFLPPTQSGYTKRFAIVDD 162  
QY 66 -----TSVDLTADGVK-----YATCQYSGEETFTSTL 94  
DB 163 AILSVGGATAPNCAQOQPPITSTNPTIDGKPMGSLPPNIEGTVYMYAGYYPM--- 218  
QY 95 TCTVNDALKSIKAFGT---VTLPIAFNVGCTGSSDLEDSKCTAGTNTVTFNDGDKD 150  
DB 219 -----KVYVSNAWSMGFLPISVTLF-----DGTIVSDDFEG-----YVYSFDD-- 256  
QY 151 ISIDVEFEKSTV-DPSAYLASKRVMSPLNKVTVLVAPOCENGYSCTGTMFSSNGDVAI 209  
DB 257 ---DLSQNCNCTVPDPSNYA-VSTTTTTEPWTGTFTSTSTMTTGTGNGVPTDEVIVI 312  
QY 210 DCSNIHIGITKGLNDWNPVSSSEFSYTKTCTNSGIIQIKYQNVNVPAGYRPFIDAYISATDV 269  
DB 313 RTPTTASTIITTEPWNSTFTSTSTELTAVTGTNGVTDFTII-----VIRTPPTATTA 366  
QY 270 NOYTLAYTNDYTCAGSLQSKPFLRWGTGKNSDAGSNGI---VIVATRTVTDSTTAVT 326  
DB 367 ITTTEPWNSTFTSTSTEL-----TTVTGTNGLPTDETIIVIRPTTATTAMT 413  
QY 327 TL-PENPSVDKTKT-----IEILOPIPTTTT-----SYGVGVTTS 361  
DB 414 TTQPNDDTFTSTSTMTTGTNGLPTDETIIVIRPTTATTAMTTPQDDTFTSTSTE 473  
QY 362 YLTKTA---PIGETATVVDVVPYHHTT--TVTSEWGTGTT-----TTTTRTN--PTDS 407  
DB 474 MITVTGTNGLPTDET-IIIVIRPTTATTAMTTPQNDFTSTSTMTTGTNGLPTD- 531  
QY 408 IDTVV 413  
DB 532 -ETIIV 536

RESULT 3  
US-08-928-361B-5  
Sequence 5, Application US/08928361B  
Patent No. 6071518  
GENERAL INFORMATION:  
APPLICANT: Petersen, Carolyn  
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
TITLE OF INVENTION: SPECIES INFECTIONS



NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-1840  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/928,361B  
FILING DATE: 12-SEP-1997  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,062  
FILING DATE: 13-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Verny, Hana  
REGISTRATION NUMBER: 30,518  
REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-324-1677  
TELEFAX: 650-324-1678  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1837 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-928-361B-5

Query Match 8.0%; Score 175.5; DB 3; Length 1837;  
Best Local Similarity 22.9%; Pred. No. 2.7e-06;  
Matches 93; Conservative 45; Mismatches 176; Indels 93; Gaps 13;  
QY 43 TSANPGDFTFLNMPCKVFKYTSQTSVDLTADGVKATCQFYSGEFTFSLTCTVNDAL 102  
DB 452 TTKKPTT 511  
QY 103 KSIKAFQTVTLPIAFNVGGTGSSTDL-----EDSKCFTAGTNTVTFNDGDKDI 151  
DB 512 PTTTTATTTT-----TTSETSVIKPDCEKCEKAGEAKGATYGVIGKDGRI 562  
QY 152 SIDVEFEKSTVPSAYLYASRVMPSLNKKVTLFVAPQCENGYSGTMGFSSNGDVAI-- 209  
DB 563 ENGMAFTMPNDOTHVRFKRVKDVGNATISV-----RCGKG--AGKLEPPDRSLDFTIPP 615  
QY 210 -----DCSNHIGITKGLNDMNPVSSESFS-----YTKTCTSNIGIQIY 249  
DB 616 VAGHNSCS-IIVGVSGGKIHVSPYSGKDVLSAPIQCELFNEVYCDCTA-----KY 669  
QY 250 QNVPGYRPFDAYISATDVNQYTLAYTNDYTCAGSRQSKPFTLRWTGYKNSDAGSNGI 309  
DB 670 GAHSGYOTSADFTVTTT-----AKPTT-----TTTGAPGQ 700  
QY 310 VIVATRTVDSSTAVTLPENPNSVDKTKTIELQPIPTTTTTSYVGVTTSYLTAKPI 369  
DB 701 PTTTTGSPSKPTTTTTT-----KATTTTLNPIITTT-TQKPTTTTITKVPKGPPI 752  
QY 370 GETAVIVDPYVTVTTTSEMGTCTTTTTTTTNTPTDSIDTVVQVP 416  
DB 753 ATTTTLAPIVTTTTTKATTTTTTTPVTTTT-TTKRDEMTTTTTPLP 798

RESULT 4  
US-08-700-651-5  
; Sequence 5, Application US/08700651B

Patent No. 6015882  
GENERAL INFORMATION:  
APPLICANT: PETERSEN, CAROLYN  
ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306-1840  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/700,651B  
FILING DATE: 08/415,751  
EARLIER FILING DATE: 1997-08-14  
EARLIER FILING DATE: 1995-04-03  
NUMBER OF SEQ ID NOS: 15  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 5  
LENGTH: 1721  
TYPE: PRT  
ORGANISM: Cryptosporidium parvum  
US-08-700-651-5

Query Match 8.0%; Score 175; DB 3; Length 1721;  
Best Local Similarity 23.2%; Pred. No. 2.7e-06;  
Matches 96; Conservative 41; Mismatches 180; Indels 96; Gaps 16;  
QY 43 TSANPGDFTFLNMPCKVFKYTSQTSVDLTADGVKATCQFYSGEFTFSLTCTVNDAL 102  
DB 326 TTTTTTTTTTKKPTT 378  
QY 103 KSIKAFQTVTL-----PIAFNVGGTGSSTDLDS-----KCFAGTNTVTF 144  
DB 379 PTTTTTTTTTTTTTKKPTT 438  
QY 145 NGDKOISIDVFEKSTVPSAYLYASRVMPSLNKKVTLFVAPQCENGYSGTMGFSSSN 204  
DB 439 IKGRIENGMAFTMPNDOTHVRFKRVKDVGNATISV-----RCRKG--AGKLEPPDRS 491  
QY 205 GDVAI-----DCS-----NTHIGI--TKGLNDMNPVS-SESFS--YTKTCTSN 243  
DB 492 LDFTIPPVAGHNSCSIIIVGVSGGKIHVSPYSGKDVLSAPIQSELFNEVYCDCTA- 550  
QY 244 GIQIKQNVPGYRPFDAYISATDVNQYTLAYTNDYTCAGSRQSKPFTLRWTGYKNSD 303  
DB 551 -----KYGAHSGYOTSADFTVTTT-----AKPTTTTTCGAPGQPT 585  
QY 304 AGSNGIVATRTVTDSTAVTLPENPNSVDKTKTIELQPIPTTTTTSYVGVTTSYL 363  
DB 586 TTTTGSPPSKPTTTTTTKATTTTTTTL--NPIITTT-----TQKPTTTTITKVPK- 631  
QY 364 TKTAPIGETAVIVDPYVTVTTTSEMGTCTTTTTTTTNTPTDSIDTVVQVP 416  
DB 632 --KPIATTTTTLKPITVTTTTTKATTTTTTTPVTTTT-TTKRDEMTTTTTPLP 681

RESULT 5  
US-08-928-361B-6  
; Sequence 6, Application US/08928361B  
Patent No. 6071518  
GENERAL INFORMATION:  
APPLICANT: Petersen, Carolyn  
TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
TITLE OF INVENTION: SPECIES INFECTIONS  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
STREET: 385 Sherman Avenue, Suite 6  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA



Qy	319	TDSTTAVTTLPNPSVDKTKTEIILOPIPTTTTTSYGVGVTSTYLTAKTAPIGETATVI-V	377
Db	350	TDETVIVIRTPTSGLISSTT---EPM-TGTFTSTEMTTVTGTNGQPTDETIVIRT	404
Qy	378	DVPYHTTTTTSVSEWGTGTTTTTR-----TN--PTDSIDTVVVQVP	416
Db	405	PTSEGLVTTTTEPWTGFTSTSTEMSTVTGTNGLPTDE-TVIVVKTP	450
<p>RESULT 7</p> <p>US-08-971-692-15</p> <p>Sequence 15, Application US/08971692</p> <p>Patent No. 6114147</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT:</p> <p>TITLE OF INVENTION: Immobilized proteins with specific binding</p> <p>TITLE OF INVENTION: Capacities and their use in processes and products.</p> <p>NUMBER OF SEQUENCES: 40</p> <p>COMPUTER READABLE FORM:</p> <p>MEDIUM TYPE: Floppy disk</p> <p>COMPUTER: IBM PC compatible</p> <p>OPERATING SYSTEM: PC-DOS/MS-DOS</p> <p>SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)</p> <p>CURRENT APPLICATION DATA:</p> <p>APPLICATION NUMBER: US/08/971,692</p> <p>FILING DATE:</p> <p>CLASSIFICATION: 435</p> <p>INFORMATION FOR SEQ. ID. NO. 15:</p> <p>SEQUENCE CHARACTERISTICS:</p> <p>LENGTH: 894 amino acids</p> <p>TYPE: amino acid</p> <p>TOPOLOGY: linear</p> <p>MOLECULE TYPE: protein</p> <p>US-08-971-692-15</p>			
<p>Query Match 7.1%; Score 155; DB 3; Length 894;</p> <p>Best Local Similarity 23.3%; Pred. No. 5.5e-05;</p> <p>Matches 109; Conservative 49; Mismatches 181; Indels 128; Gaps 22;</p>			
Qy	12	NSUTSNAANAFAKPGPGYPTWNAVGLSLDGTSANPGDTFTLNNPCV-----	58
Db	50	DSSTYSNAAMAY---GYASKT-----KLAGSVGGQDTISIDYNIPCVSSSGTFFCPQEDS	101
Qy	59	-----FKYTTQSOTSDLTADGVKATQCYSGEFTTFTST	93
Db	102	YGMNCKGMGACSNQGIAYWSTDLFGFYTPPTNVLTMTG-----YFLPQTGTSYTF	154
Qy	94	LTCVTNDALKSSIKAFQTVTLPIAFNVGGTGSSTDLSDSKCTAGTNTVTFNDGDKDI--	151
Db	155	KFATVDD---SAILSVG-----GATAFCCAAQOQPPITSTNFTI---DGILKPMWG	198
Qy	152	SIDVEFEKSTVDPDSAYLYASRVMPSLNKVT--TLFVAPQCENGYT-----SGTMG	199
Db	199	SLPPNIEGTVMYAGYYPKVYVS-NAVSMGTLPISVTLPDGTVYSDDFEGVYVSDDD	257
Qy	200	FSSNGDVAICDSNIHIGITKGLND-4NYPVSSSESFTKCTNSGIIQIKYQNPAGYRP	258
Db	258	LSOSNCTVP-DPSNVAVSITTTTTTPEWTGFTSTEMTTVTGTNGQPTD-ETVIVIRTP	315
Qy	259	FIDAYISATDVNQVTLAYNDVTCAGSLKOSKPFLLURVTGYKNSDAGSNGIVIVATRTV	318
Db	316	TSEGLISSTT-----TEPMTGFTTS--TSTEVTTTGTNGQP	349
Qy	319	TDSTTAVTTLPNPSVDKTKTEIILOPIPTTTTTSYGVGVTSTYLTAKTAPIGETATVI-V	377
Db	350	TDETVIVIRTPTSGLISSTT---EPM-TGTFTSTEMTTVTGTNGQPTDETIVIRT	404
Qy	378	DVPYHTTTTTSVSEWGTGTTTTTR-----TN--PTDSIDTVVVQVP	416
Db	405	PTSEGLVTTTTEPWTGFTSTSTEMSTVTGTNGLPTDE-TVIVVKTP	450

```

: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 216 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-928-361B-8

Query Match          6.8%  Score 148.5, DB 3; Length 216;
Best Local Similarity 31.5%  Pred. No. 2.4e-05;
Matches 46; Conservative 10; Mismatches 79; Indels 11; Gaps

QY 265 SATVDNYUVLAVYNDYTCAGSRQSKPFLRWYGYKNSDAGSGIVVATRTFTVDSTTA 324
Db 44 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTT-----TTTTTTTTTTTTTTTT 95
QY 325 VTLLPFPNSVDKTKLIELQIPFTTITTSYGVGVTTSVLTKTAPIGETATVIVDVPVHTT 384
Db 96 TTTTNTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTTT-----TTTTTTTTTTTTTTTT 152
QY 385 TTVTSEWGTGTTTTTTRTNPTDSIDT 410
Db 153 TTTTNTTTTTTTTTTTTTTTTTTTTTTTT 178

RESULT 11
US-08-325-367A-4
: Sequence 4, Application US/08325267A
: Patent No. 5585271
: GENERAL INFORMATION:
: APPLICANT: WATARI, JUNJI
: APPLICANT: TAKATA, YOSHIHIRO
: APPLICANT: OGAWA, MASAHIRO
: APPLICANT: PENTILLA, MERJA
: APPLICANT: ONNELA, MAIJA-LEENA
: APPLICANT: KERANEN, SIRKA
: TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT
: STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400
: CITY: ARLINGTON
: STATE: VA
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/325,267A
: FILING DATE: 18-NOV-1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP PCT/JP94/00290
: FILING DATE: 24-FEB-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 38871/1993
: FILING DATE: 26-FEB-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: OBLON, NORMAN F
: REGISTRATION NUMBER: 24,618
: REFERENCE/DOCKET NUMBER: 2589-023-0XPCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-413-3000
: TELEFAX: 703-413-2220
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:

```

;  
; LENGTH: 862 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-325-267A-4

Query Match 6.8%; Score 148.5; DB 1; Length 862;  
Best Local Similarity 23.3%; Pred. No. 0.00019;  
Matches 107; Conservative 45; Mismatches 149; Indels 159; Gaps 23;

QY 12 NSLTWSNAANYAFKGPYPTWNAVLGMSLDGTSANPGDFTFLNMPCKVFKYTTTSQTSVDLT 71  
DB 50 DSYTYSNAYMAY---GVASKT-----KLGSGGGQTDISIDYNIPIV---SSSGTTPCPQ 98  
QY 72 ADGVKATQO-----FVSGEETTFSTLCTVNDALKSSIKAFGTVTLPIA-- 117  
DB 99 EDSYGNMGCKGACNSOGIAYNSTDLFGYTTPT-----NVTLEMTGY 143  
QY 118 FNVGGTSSDLEDSKCFCTAGTNTVTFNDGDKDISIDVEFEKSTVDPSSAVLYASRWPSL 177  
DB 144 FLPQTGSYT-----FKFATVDDSAIL-----SV 167  
QY 178 NKVITLFAPOCENGYSCTMGSSNGDVAIDCSNHHGKGLNDW--NYPVSSESES 235  
DB 168 GGATAPNCAQQOPIITS-----TNFTI---DGKPHGSLPNIEGTV 208  
QY 236 YTKTCTSGNGIQIKYON-VPAGYRPFI---DAYISATDVNOYTLAYND----- 279  
DB 209 IYAGYYPKMYVSNVSNVSMGTLPISVLPDGTVSDDFEGYVSFDDLSQSNCTVDPDP 268  
QY 280 --YTCAGSRLOSKPRLTWGKYNSDA-----GNGI-----VIVATRTVTDSTAV 325  
DB 269 SNYAVSTTTTTEP-----WGTFTSTSTEMTVTGNGVPTDETIVIRPT-TASTIIT 323  
QY 326 TLLFPNPSVDKTKTIEILOIPIPTTTTSYGVGVTTSYLNKTAIPIGETATVI-VDVPYHTT 384  
DB 324 TTEPM-----TGFTSTSTEMTVTGNGQPTDETIVIRPTSEGLV 366  
QY 385 TTVTSEWTCGTTTTTR-----TN--PTDSIDTVVYVQVP 416  
DB 367 TTTTPTWGTGTSTSTEMTVTGTLNGLPTDE-TVIVYVKT 405

RESULT 12  
US-08-700-651-15  
; Sequence 15, Application US/08700651B  
; Patent No. 6015882  
; GENERAL INFORMATION:  
; APPLICANT: PETERSEN, CAROLYN  
; APPLICANT: LEECH, JAMES  
; APPLICANT: NELSON, RICHARD, C.  
; APPLICANT: GUT, JIRI  
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS  
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum  
; FILE REFERENCE: 480.19-4(HV)  
; CURRENT FILING DATE: 1997-08-14  
; EARLIER FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 15  
; LENGTH: 249  
; TYPE: PRT  
; ORGANISM: Cryptosporidium parvum  
; FEATURE:  
; OTHER INFORMATION: mutant/variant of SEQ ID NO:5  
US-08-700-651-15

Query Match 6.7%; Score 146.5; DB 3; Length 249;

Best Local Similarity 33.3%; Pred. No. 4.4e-05;  
Matches 40; Conservative 11; Mismatches 68; Indels 1; Gaps 1;  
QY 291 PFTLRWTKYKNSDAGSNGIVIVATRTVTDSTAVTTLFPNPSVDKTKTIEILQPIPTTT 350  
DB 8 PYT-KCVGKHTT 66  
QY 351 ITTSYGVGVTTSYLNKTAIPIGETATVIVDPVPHHTTTTIVTSEWTCGTTTTTTRNPTDSIDT 410  
DB 67 TT 126

RESULT 13  
US-08-928-361B-20  
; Sequence 20, Application US/08928361B  
; Patent No. 6071518  
; GENERAL INFORMATION:  
; APPLICANT: Petersen, Carolyn  
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,  
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS  
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM  
; TITLE OF INVENTION: SPECIES INFECTIONS  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA  
; STREET: 385 Sherman Avenue, Suite 6  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306-1840  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/928,361B  
; FILING DATE: 12-SEP-1997  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/026,062  
; FILING DATE: 13-SEP-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Verny, Hana  
; REGISTRATION NUMBER: 30,518  
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)  
; TELEPHONE: 650-324-1677  
; TELEFAX: 650-324-1677  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 249 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-928-361B-20

Query Match 6.7%; Score 146.5; DB 3; Length 249;  
Best Local Similarity 33.3%; Pred. No. 4.4e-05;  
Matches 40; Conservative 11; Mismatches 68; Indels 1; Gaps 1;

QY 291 PFTLRWTKYKNSDAGSNGIVIVATRTVTDSTAVTTLFPNPSVDKTKTIEILQPIPTTT 350  
DB 8 PYT-KCVGKHTT 66  
QY 351 ITTSYGVGVTTSYLNKTAIPIGETATVIVDPVPHHTTTTIVTSEWTCGTTTTTTRNPTDSIDT 410  
DB 67 TT 126

RESULT 14

```
US-08-700-651-9
; Sequence 9, Application US/08700651B
; Patent No. 6015882
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: VACCINES, ANTIBODIES, PROTEINS, GLYCOPROTEINS, DNAS AND RNAS
; TITLE OF INVENTION: FOR PROPHYLAXIS AND TREATMENT OF Cryptosporidium parvum
; TITLE OF INVENTION: INFECTIONS
; FILE REFERENCE: 480.19-4(HV)
; CURRENT APPLICATION NUMBER: US/08/700,651B
; CURRENT FILING DATE: 1997-08-14
; EARLIER APPLICATION NUMBER: 08/415,751
; EARLIER FILING DATE: 1995-04-03
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 130
; TYPE: PRT
; ORGANISM: Cryptosporidium parvum
; FEATURE:
; OTHER INFORMATION: mutant/variant of SEQ ID NO:5
US-08-700-651-9

Query Match          6.68; Score 145; DB 3; Length 130;
Best Local Similarity 33.38; Pred. No. 2.2e-05;
Matches 42; Conservative 12; Mismatches 62; Indels 10; Gaps 2;

QY 291 PFTLRWTGYKNSDAGSNGIVAVTRTVDSTAVTTLPEPNPSVDKTKTIEILOPIPTTT 350
Db 8 PYT-KCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 57

QY 351 ITTSYGVGVTTSYLTAPIGETATVIVDPYHTTTTSEMTCTITTTTTRNPTDSIDT 410
Db 58 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTATTTTSET 117

QY 411 VVQVVP 416
Db 118 ESVIKP 123

RESULT 15
US-08-928-361B-14
; Sequence 14, Application US/08928361B
; Patent No. 6071518
; GENERAL INFORMATION:
; APPLICANT: Petersen, Carolyn
; TITLE OF INVENTION: PEPTIDES, POLYPEPTIDES, GLYCOPROTEINS,
; TITLE OF INVENTION: THEIR FUNCTIONAL MUTANTS, VARIANTS, ANALOGS AND FRAGMENTS
; TITLE OF INVENTION: FOR TREATMENT AND DETECTION/DIAGNOSIS OF CRYPTOSPORIDIUM
; TITLE OF INVENTION: SPECIES INFECTIONS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PETERS, VERNY, JONES & BIKSA
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,361B
; FILING DATE: 12-SEP-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/026,062
```

```
; FILING DATE: 13-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Verny, Hana
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.76-1(HV)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-1677
; TELEFAX: 650-324-1678
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 130 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-928-361B-14

Query Match          6.68; Score 145; DB 3; Length 130;
Best Local Similarity 33.38; Pred. No. 2.2e-05;
Matches 42; Conservative 12; Mismatches 62; Indels 10; Gaps 2;

QY 291 PFTLRWTGYKNSDAGSNGIVAVTRTVDSTAVTTLPEPNPSVDKTKTIEILOPIPTTT 350
Db 8 PYT-KCVGVKHTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 57

QY 351 ITTSYGVGVTTSYLTAPIGETATVIVDPYHTTTTSEMTCTITTTTTRNPTDSIDT 410
Db 58 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTKKPTTTTATTTTSET 117

QY 411 VVQVVP 416
Db 118 ESVIKP 123

Search completed: October 3, 2002, 16:30:03
Job time: 151 sec
```



**THIS PAGE BLANK (USP)**



Fri Oct 4 08:13:54 2002

us-09-715-876-8\_copy\_17\_432.text.rpr

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:28:17 ; Search time 20.85 Seconds  
(without alignments)  
1917.178 Million cell updates/sec

Title: US-09-715-876-8\_COPY\_17\_432

Perfect score: 2190

Sequence: 1 AKTTTGVDFSNLWSNAA.....TTTTTNTPTDSIDTVVQVP 416

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR71.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2190	100.0	1260	2 S60896	agglutinin-like pr
2	1907	87.1	1419	2 T30531	agglutinin-like ad
3	308.5	14.1	650	2 S22835	alpha-agglutinin -
4	198.5	9.1	948	2 T11678	hypothetical prote
5	177	8.1	1367	2 S51959	hypothetical prote
6	177	8.1	1537	2 S53465	flocculation prote
7	175.5	8.0	1832	2 T31113	mucin-like glycopr
8	174.5	8.0	973	2 T40778	hypothetical 129.5
9	171.5	7.8	456	2 T21389	hypothetical serin
10	171	7.8	1251	2 T38221	hypothetical prote
11	167.5	7.6	662	2 A45155	mucin RIM-C.1 - Af
12	161.5	7.4	1075	2 S48992	flocculation prote
13	161	7.4	1428	2 AC2224	hypothetical prote
14	156.5	7.1	770	2 T22808	hypothetical prote
15	156	7.1	851	2 T22696	hypothetical prote
16	155	7.1	597	2 B75556	hypothetical prote
17	155	7.1	1180	2 B86719	hypothetical prote
18	154.5	7.1	1367	1 S48476	hypothetical prote
19	154	7.0	1283	2 T39174	glucan 1,4-alpha-g
20	153.5	7.0	3570	2 T45258	hypothetical Serin
21	153.5	7.0	792	2 T45035	a-agglutinin core
22	153	7.0	792	2 T50305	mucin MUC5B, trach
23	152	6.9	825	2 T29634	hypothetical prote
24	152	6.9	1902	2 B45764	hypothetical prote
25	151.5	6.9	1797	2 T21889	lactococpin (EC 3.4
26	151.5	6.9	1805	2 T21888	hypothetical prote
27	148.5	6.8	1169	2 S38181	hypothetical prote
28	145.5	6.6	648	2 PC4395	flocculation prote
29	145	6.6	995	2 S50358	mucin 3 - human (f
					hypothetical prote

30	145	6.6	1041	2 S55862	pro
31	145	6.6	1609	2 S25345	pro
32	145	6.6	1962	2 A32634	lact
33	144.5	6.6	609	2 S62518	hypot
34	144.5	6.6	881	2 S56032	probab
35	143.5	6.6	3029	2 S76109	hypothet
36	143	6.5	798	2 T34248	hypothet
37	143	6.5	1902	2 S06997	lactococpin EC 3.4
38	142.5	6.5	1275	2 T33369	hypothetical prote
39	142	6.5	522	2 S41819	nucleoporin p62 -
40	142	6.5	583	2 S67571	hypothetical prote
41	142	6.5	1777	2 T34369	hypothetical prote
42	141.5	6.5	967	2 S66852	hypothetical prote
43	141	6.4	1902	1 B44858	lactococpin (EC 3.4
44	140.5	6.4	691	2 B75622	hypothetical prote
45	140	6.4	5291	2 F30636	hypothetical prote

#### ALIGNMENTS

RESULT 1  
S60896  
agglutinin-like protein - yeast (Candida albicans)  
C:Species: Candida albicans  
C:Date: 27-Apr-1996 #sequence\_revision 13-Mar-1997 #text\_change 17-Mar-2000  
C:Accession: S60896  
R:Hoeyer, L.L.; Scherer, S.; Shatzman, A.R.; Livvi, G.P.  
Mol. Microbiol. 15, 39-54, 1995  
A:Title: Candida albicans ALS1: domains related to a Saccharomyces cerevisiae sexual  
A:Reference number: S60896; MUID:95272392  
A:Accession: S60896  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1260 <HOY>  
A:Cross-references: EMBL:L25902; NID:g704426; PIDN:AAC1649.1; PID:g704427  
C:Superfamily: Yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida

Query Match 100.0%; Score 2190; DB 2; Length 1260;  
Best Local Similarity 100.0%; Pred. No. 9.2e-136;  
Matches 416; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AKTTTGVDFSNLWSNAA	YAFKPGYPTWNAVLGWSLDGTSANPGDTFTLNMPGVFK	60
DB	17	AKTTTGVDFSNLWSNAA	YAFKPGYPTWNAVLGWSLDGTSANPGDTFTLNMPGVFK	76
QY	61	YTTTSQTSVLTADGVKYAT	COFYSGEEFTTFTLTCTVNDALKSSIKAFGTVTLPIAFNV	120
DB	77	YTTTSQTSVLTADGVKYAT	COFYSGEEFTTFTLTCTVNDALKSSIKAFGTVTLPIAFNV	136
QY	121	GGTGSSTDLDSKCF	TAGTNTVTFNDGDKDISIDVEFEKSTVDPSSAYLYASRVMPSLNKV	180
DB	137	GGTGSSTDLDSKCF	TAGTNTVTFNDGDKDISIDVEFEKSTVDPSSAYLYASRVMPSLNKV	196
QY	181	TLFLVAPQCE	NGYSGTWMGSSNGDVAIDCSNIHIGITKGLNDWNPVSSSEFSYTKC	240
DB	197	TLFLVAPQCE	NGYSGTWMGSSNGDVAIDCSNIHIGITKGLNDWNPVSSSEFSYTKC	256
QY	241	TSNGIQIKQNP	PAGYRPFIDATISATDNYQTYLATYNDYTCAGSRQSKPFTLRWTGK	300
DB	257	TSNGIQIKQNP	PAGYRPFIDATISATDNYQTYLATYNDYTCAGSRQSKPFTLRWTGK	316
QY	301	NSDAGSNG	IVVATRTVTDSTAVTTLFPNPSVDKTKTIEILOPIPTTTTTSYVGVT	360
DB	317	NSDAGSNG	IVVATRTVTDSTAVTTLFPNPSVDKTKTIEILOPIPTTTTTSYVGVT	376
QY	361	SYLTKTAP	IGETATVIVDVPIHTTIVTSEWGTGTTTTTTRNPTDSIDTVVQVP	416
DB	377	SYLTKTAP	IGETATVIVDVPIHTTIVTSEWGTGTTTTTTRNPTDSIDTVVQVP	432

RESULT 2

T30531  
agglutinin-like adhesin - yeast (Candida albicans)  
C:Species: Candida albicans  
C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 17-Mar-2000  
C:Accession: T30531  
R:Gaur, N.K.; Klotz, S.A.  
Infect. Immun. 65, 5289-5294, 1997  
A:Title: Expression, cloning, and characterization of a Candida albicans gene, ALA1, the  
A:Reference number: Z20847; MUID:98053977  
A:Accession: T30531  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1419 <GAU>  
A:Cross-references: EMBL:AF025429; NID:g2522218; PID:g2522219; PIDN:AAB88883.1  
C:Genetics:  
A:Gene: ALA1  
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

Query Match 87.18; Score 1907; DB 2; Length 1419;  
Best Local Similarity 86.34; Pred. No. 3.9e-117;  
Matches 359; Conservative 21; Mismatches 36; Indels 0; Gaps 0;

QY 1 AKTGVDFSNLSWNAAYAKGPGYTNWNLGSLDGTSAHPGDTFLNMPGVFK 60  
DB 17 AKATGIFNSIDSLWNAAGNAGKPGYTNWNLGSLDGTSAHPGDTFLNMPGVFK 76  
QY 61 YTTQTSVDLTADGVKATCFQYSGEFTTFLCTCTVNDALKSSIAKAGTGTTLPIAFNV 120  
DB 77 FTASOKSVDLTADGVKATCFQYSGEFTTFLCTCTVNNLRSSIKALGTGTTLPIAFNV 136  
QY 121 GGTGSSDLEDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDPSPAYLYASRMPSLNKV 180  
DB 137 GGTGSSVDLEDSKCFAGTNTVTFNDGSKLSTAVNFESTVDQSGYLTTSRPMPSLNKI 196  
QY 181 TFLVAPQCENGYTSCTMGFSSNGDAIDCSNHHIGITKGLNDWNPVSSFSYKTC 240  
DB 197 ATLVAPQCENGYTSCTMGFSTGYDVAIDCSNVHIGISKVDNHPVTSFSYKTC 256  
QY 241 TSNQIKYQNVAGYRPFIDAYISATDVNQVTLAYTNDYTCAGSLQSKPFTLRWTCYK 300  
DB 257 SFSGISITYQNVAGYRPFIDAYISPDNNQVQLSYKNDYTCDDVYQWQAPFTLRWTCYK 316  
QY 301 NSDAGSNGIVAVTRTVDSTATTTLTPNPSVDKTKTIEIQLPIPTTTITTSYGVVTT 360  
DB 317 NSDAGSNGIVAVTRTVDSTATTTLTPNPSVDKTKTIEIQLPIPTTTITTSYGVVTT 376  
361 SYLTKTAPIGATATVVDVPYHTTTTTSMTGTITTTTTRTNPDSIDTVVQVP 416  
DB 377 SYSTKAPIGATATVVDVPYHTTTTTSMTGTITTTTTRTNPDSIDTVVQVP 432

RESULT 3  
S22835  
alpha-agglutinin - yeast (Saccharomyces cerevisiae)  
N:Alternate names: 22K glycoprotein; protein J1418; protein YJR004c  
C:Species: Saccharomyces cerevisiae  
C>Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 29-Oct-1999  
C:Accession: S22835; S51229; A32822; S55192; S57019  
R:Hauser, K.; Tanner, W.  
FEBS Lett. 255, 290-294, 1989  
A:Title: Purification of the inducible alpha-agglutinin of S. cerevisiae and molecular  
A:Reference number: S22835; MUID:50005993  
A:Accession: S22835  
A:Molecule type: DNA  
A:Residues: 1-650 <HAU>  
A:Cross-references: EMBL:X16861; NID:g3352; PIDN:CAA34752.1; PID:g3353  
A:Accession: S51229  
A:Molecule type: Protein  
A:Residues: 20-24 <RAZ>  
R:Li, P. N.; Wojcikowicz, D.; Kurjan, J.  
Mol. Cell. Biol. 9, 3155-3165, 1989  
A:Title: AG-alpha-1 is the structural gene for the Saccharomyces cerevisiae alpha-agglut

A:Reference number: A32822; MUID:90014768  
A:Accession: A32822  
A:Molecule type: DNA  
A:Residues: 1-448, 'P', 450-555, 'E', 557-580, 'L', 582-650 <LIP>  
A:Cross-references: GB:M28164; NID:g171041; PIDN:AAA34417.1; PID:g171044  
R:de Haan, M.; Smits, P. H. M.; Grivell, L. A.  
submitted to the EMBL Data Library, May 1995  
A:Reference number: S55183  
A:Accession: S55192  
A:Molecule type: DNA  
A:Residues: 1-650 <DEH>  
A:Cross-references: EMBL:X87611; NID:g854567; PIDN:CAA60926.1; PID:g854577  
R:de Haan, M.; Grivell, L. A.; Smits, P. H. M.  
submitted to the Protein Sequence Database, September 1995  
A:Reference number: S56771  
A:Accession: S57019  
A:Molecule type: DNA  
A:Residues: 1-650 <ZAG>  
A:Cross-references: EMBL:Z49504; NID:g1015625; PIDN:CAA89526.1; PID:g1015626; MIPS:YJ  
C:Genetics:  
A:Gene: SGD:SAG1; AGALL; AGALPHA1  
A:Cross-references: SGD:S0003764; MIPS:YJR004c  
A:Map position: 10R  
C:Keywords: glycoprotein

Query Match 14.18; Score 308.5; DB 2; Length 650;  
Best Local Similarity 25.88; Pred. No. 9.9e-13;  
Matches 116; Conservative 79; Mismatches 190; Indels 65; Gaps 20;

QY 10 SFNSLWSN-----AANYAFKPGYPTNAVGLWSL-DGTSANPGDTFLNMPGVFK-- 60  
DB 21 NINDITFSNLEITPLTAN---KQPD-OGMTATDFSDIADASSIREGDEFTLSMPHYRIK 76  
QY 61 -YTTQTSVDLTADGVKATCFQYSGE-----EFTTSTLTCTVNDALKSSIKAFGVTL 114  
DB 77 LLNSSTATISLADTEAFKC-YVSQAAAYLYENTTF---TCTAQNLSYNTIDGSIITF 132  
QY 115 PIAPNVGGTGSSTDLSDSKOFTAGTNTVTFNDGDKDISIDVEFEKSTVDPSPAY---LYAS 171  
DB 133 SLNFSGGSGSYELENNAKFKSGPMLVKLGQMSDV---VNF-----DPAFTENVFHS 184  
QY 172 RVMPSLNKVTTFLVAPQCENGYTSG---TMGFSSNGDVAIDCSNHHIGITKGLNDWNP 228  
DB 185 GRSTGYGSFESYHLGMYCPNGYFLGGTEKIDYDSSNNNDVDCSSVQVYSSNDFNDWMP 244  
QY 229 VSSEFSYTKTCTCTNGIQIKY-QNVAGYRPFIDAYIS-ATDVNQVTLAYTNDYTC---- 282  
DB 245 QSYNDTNAVDTCTGSLNWLITLDEKLYDGEMLVNALQSLPANVNTIDHALEFQYTCLODI 304  
QY 283 AGSLQSKPFTLR---WTGYKNSDAGSNGIVAVTRTVD-----STTAVTTLFPNP 332  
DB 305 ANTYATQFSTTREFIVYQGNLGTASAKSSFTSTTTDLTSTINTSAYSTGSIIV- 360  
QY 333 SVDTKTITILQIPIPTTTITTSYGVVTSYLTAPIGETATVVDVPYHTTT----- 385  
DB 361 ETGNRTTSEVISHVVTSTKLSPATTATSLTIAQTSIYSDTSNITVGTDIHTTSEVSDVE 420  
QY 386 TVTSEMTGTITTT-TTRTNPDSIDTVVQV 414  
DB 421 TISRETASTVAAPTSTTGTGAMNTYISQ 450

RESULT 4  
T11678  
hypothetical protein SPBC21D10.06c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C:Accession: T11678  
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M. A.; Barrell, B. G.  
submitted to the EMBL Data Library, September 1998  
A:Reference number: Z17313  
A:Accession: T11678

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-948 <SEE>

A:Cross-references: EMBL:AL031536; NID:e1319499; PID:e1319505

A:Experimental source: strain 972h(-)

C:Genetics:

A:Map position: IIR

A:Note: SPBC21D10.06c

Query Match 9.1%; Score 198.5; DB 2; Length 948;  
Best Local Similarity 23.4%; Pred. No. 2.6e+05;  
Matches 100; Conservative 74; Mismatches 133; Indels 121; Gaps 22;

QY 9 DFNLSLWNAANYAFKPGYPTWNAVLQWSLDGTSANPGDFTLNMPCVEKYTTSTSV 68

DB 468 NFSSISGSSASSLPITPSSVLS-NTTLHSSVOGSQS--QFTSVSPS----- 512

QY 69 DLTADGVKATCOFYSGBEFTFTLTCTVNDALKSSIKAFGTVLPFAFNVGCTGS-- 125

DB 513 -----STOSYSTSSNFTPI--TISTLSSEFPTIVSSSFQYSSLSNVT 555

QY 126 STDLSDSKCTAGTNTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKVTLFV 185

DB 556 TTAAGSSLSNSSNNAI-----HISSVISGSS--SSALSSSTIVSSINSSSSVFI 605

QY 186 AP-----OCENG-----TSTMGCFSSNGVDADCSNIHIGITKGLNDWNPVSSSESFS 235

DB 606 SSVSSLOYSSSVIETTTSGVGFTI--IATP-----VGTAGTVVVDIPTPS--W 654

QY 236 YKCTNSGQIKYQNPAGYRPIDAYTSARDVQNTLAYTNDYTCAGSRLOSKPFTLR 295

DB 655 VTEVTSGSV-----GFTTTIATPVGST-----AGTVLVDIP-TPS 689

QY 296 WTGYKNSDAGSNGIVATRTVTDSTAVTLPNPSVDKTKTIEILOPIP---TTTTI 352

DB 690 W-----VTEVTSVSGVEFTTTIATPVGTAGTVVVDIPTPSMWVETVT 732

QY 353 TSYGVVTSYLTAKTAPIGETA-TVIVDVVPHY--TTVTSEMTGIIITTTTRNPT-DSI 408

DB 733 SSGVFTTTIAT---PIGTAGTVLVDIPTPSMWVETVTS---GSVGTFTTIATPVGTGA 786

QY 409 DRVVVQVP 416

DB 787 GTVLVDIP 794

RESULT 5

S51959

hypothetical protein YAL063c - yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 22-Oct-1999

C:Accession: S51959

R:Bussey, H.; Kaback, D.B.; Zhong, W.; Vo, D.T.; Clark, M.W.; Fortin, N.; Hall, J.; Ouel

submitted to the EMBL Data Library, August 1994

A:Description: The sequence of chromosome 1 of Saccharomyces cerevisiae.

A:Reference number: S51956

A:Molecule type: DNA

A:Residues: 1-1367 <BUS>

A:Cross-references: EMBL:U12980; GSPDB:GN00001; MIPS:YAL063c

C:Genetics:

A:Gene position: 1L

Query Match 8.1%; Score 177; DB 2; Length 1367;

Best Local Similarity 23.3%; Pred. No. 0.0011;

Matches 112; Conservative 53; Mismatches 181; Indels 134; Gaps 26;

QY 21 NYAFKPGYPTWNAVLG-WSLD--GTSANPGDFTLNM-----PCVFKYTTSTSV 67

DB 104 NMKGKIGACSNPIIAYWSTDLGFGFTTPTNV-TLEMTGYFLPPQGTGYTFKATVDDS 162

QY 68 VDLTADG-VKATCOFYSGBEFTFTLTCTVNDALKSSIKAF-----GTV----- 112

DB 163 AILSVGSGSIAPFCC-----AQEQPITSTNFTIN-----GIKPMGSPPDNITCTVVMYAG 213

QY 113 -----TLPIAFNV-GGTGSGSTDLSDSKCTAGTNTVTFNDGDKDISIDV 155

DB 214 FYYPKIVYSNVAVAGTLPISVTLPGDITVSDDFEG-----YVVTFF-----DNNLS--- 259

QY 156 EFEKSTVDPDSAYLYASRVMPSLNKVTLFVAQCENGYSCTMGFSSNGDVAIDCSNIH 215

DB 260 QPNCTIPDPSNVT-VSTTIITTEPMTGTFTSTSTEMTVTGTNGVPTDETIVIRPTTA 318

QY 216 IGITKGLNDWNPVSSSESFSYTKTCTNSGIOIKYQNPAGYRPIDAYISATDVNOYTLA 275

DB 319 STIITTEPWNSTFTSTSTELTTVTGTNGVTRDETII-----VIRPTTATTATITTEP 372

QY 276 YTNDYTCAGSRLOSKPFTLRWTKYKNSDAGSNGI---VIVATRTVTDSTTAVTTL-PEN 331

DB 373 WNSTFTSTSTEL-----TTVTGTNGLPTDETIIVIRPTTATTATTTPQPN 419

QY 332 PSVDKTKT-----IEILOPIPTTITT-----SVGVVTSYLTAKTA 367

DB 420 DTFSTSTELTTVTGTNGLPTDETIIVIRPTTATTATTTPQNDFTSTSTELTTVTG 479

QY 368 ----PGETATVIVDVVPHYTTT--TVTSEWGTIT-----TTTTTRN--PTDSIDTVVV 413

DB 480 TNGLPTDET-IIVIRPTTATTATTTPQNDFTSTSTELTTVTGTNGLPTD--ETIIV 536

RESULT 6

S53465

flocculation protein FLOI precursor - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YAR050w

C:Species: Saccharomyces cerevisiae

C:Date: 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 29-Oct-1999

C:Accession: S53465; S43543; S57851; S31230; S57349

R:Bussey, H.; Keng, T.; Storms, R.K.; Vo, D.; Zhong, W.; Fortin, N.; Barton, A.B.; Ka

submitted to the EMBL Data Library, February 1994

A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: analysis of th

A:Reference number: S53458

A:Molecule type: DNA

A:Residues: 1-1537 <BUS>

A:Cross-references: EMBL:L28920; NID:gl616966; PID:AAO9499.1; PID:g694125; MIPS:YAR

R:Watari, J.; Takata, Y.; Ogawa, M.; Sahara, H.; Koshino, S.; Onnela, M.L.; Alrksaine

Yeast 10, 211-225, 1994

A:Title: Molecular cloning and analysis of the yeast flocculation gene FLOI.

A:Reference number: S43543; MUID:94262325

A:Molecule type: DNA

A:Residues: 1-428, 'M', 430-473, 'M', 520-549, 'T', 551-608, 'L', 610-636, 'M', 638

A:Cross-references: EMBL:X78160

R:Onnela, M.

submitted to the EMBL Data Library, September 1994

A:Reference number: S57851

A:Molecule type: DNA

A:Residues: 1-428, 'M', 430-463, 'D', 465-473, 'M', 475-518, 'M', 520-549, 'T', 551-608, 'L', 610

A:Cross-references: EMBL:X78160; NID:g535933; PID:CAAS5024.1; PID:g535934

R:Teunissen, A.W.R.H.; Holub, E.; van der Hucht, J.; van den Berg, J.A.; Steensma, H

Yeast 9, 423-427, 1993

A:Title: Sequence of the open reading frame of the FLOI gene from Saccharomyces cerev

A:Reference number: S31230; MUID:93289821

A:Accession: S31230

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-296, 927-1516, 'TAYWVVV' <TEU>

R:Bizard, F.; Bony, M.; Blondin, B.; Dequin, S.; Barre, P.

Yeast 11, 809-822, 1995

A:Title: The Saccharomyces cerevisiae FLOI flocculation gene encodes for a cell surfa

A:Reference number: S57349; MUID:96090130

A:Accession: S57349

A:Molecule type: DNA  
A:Residues: 1243-1274;1308-1339;1359-1390 <RID>  
C:Genetics:  
A:Gene: SGD:FLO1  
A:Map position: 1R  
A:Cross-references: SGD:S0000084; MIPS:VAR050W  
C:Keywords: duplication; glycoprotein; transmembrane protein  
F:1-24/Domain: signal sequence #status predicted <SIG>  
F:25-1537/Product: flocculation protein FLO1 #status predicted <MAT>  
F:953-997/Domain: repeat A2 <RA2>  
F:998-1042/Domain: repeat A3 <RA3>  
F:1043-1081/Domain: repeat A4 <RA4>  
F:1226-1276/Domain: repeat B1 <RB1>  
F:1277-1284/Domain: repeat B2 <RB2>  
F:1291-1341/Domain: repeat B3 <RB3>  
F:1342-1392/Domain: repeat B4 <RB4>  
F:1408-1416/Domain: repeat C1 <RC1>  
F:1417-1425/Domain: repeat C2 <RC2>  
F:1426-1434/Domain: repeat C3 <RC3>  
135-187,262,1114/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 8.1%; Score 177; DB 2; Length 1537;  
Best Local Similarity 23.0%; Pred. No. 0.0013;  
Matches 112; Conservative 47; Mismatches 181; Indels 146; Gaps 25;

Qy 21 NYAFKGGC-IPTNNVGLWSLD--GTSANPGDFTFLNM-----PCVFYKTSQ-- 65  
Db 104 NMCGMGACNSQGIAYWSTDLFGFYTPNV-TLEMTGYLPPQTGSGYTFKFAVDSD 162  
Qy 66 -----TSVDLRADGVK-----YATCQPSGEEFTFTSL 94  
Db 163 AILSVGGATAFNCAQQPITSTNFTIDGKPMGSLPNIEGVVMAYIYPM---- 218  
Qy 95 TCTVNDALKSSIKAFGT----VTLPFANVGGTSGSTDLSDSKCFAGTNVTFNDGDKD 150  
Db 219 -----KVYISNAVSWGTLPISVTLF-----DGTVSDDFEG-----YVYFDD-- 256  
Qy 151 ISIDVEFEKST-DPSAYLIASRVMSLNVKVTTLFVAPOCENGYSGTMGFSSSNGDAI 209  
Db 257 ---DLSQSNCTVDPDSNYA-VSTTTTTEPMTGTFTSTEMTTVTGTNGVPVDETVI 312  
Qy 210 DCSNIHIGITKGLNDWNPVSSSESVKTKTSNGIOIKYQNVAGYRPFIDAYISATDV 269  
Db 313 RPTPTASTIITTEPWNSTFTSTSTELTGTGTNGVTRDTETII-----VIRPTTATTA 366  
Qy 270 NOYTLATNDYTCAGSLQSKPFTLRMTGKNSDAGSNGI---VIVATRTVTDSTAVT 326  
Db 367 ITTTEPWNSTFTSTSTEL-----TTVTGNGLPDDETIIIVIRPTTATTA 413  
Qy 327 TL-PNPSVDKTKT-----IEILOPIPTTTIT-----SYVGVVTS 361  
Db 414 TTOPNDFTSTSTELTGTGTNGLPDDETIIIVIRPTTATTA 473  
Qy 362 YLTAKTA---PIGETATVVDVVPYHTTT--TVTSEWGTIT-----TTTTRTN--PTDS 407  
Db 474 LTTVTGTNGLPDDET-IIIVIRPTTATTA 531  
Qy 408 IDTVVV 413  
Db 532 -ETIIV 536

RESULT 7  
mucin-like glycoprotein 900 - Cryptosporidium parvum  
C:Species: Cryptosporidium parvum  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 22-Oct-1999  
R:Accession: T31113  
R:Barres, D.A.; Bonnin, A.; Huang, J.X.; Gousset, L.; Wu, J.; Gut, J.; Dubrem  
Mol. Biochem. Parasitol. 96, 93-110, 1998  
A:Title: A novel multi-domain mucin-like glycoprotein of Cryptosporidium parvum mediates  
A:Reference number: Z20989; MUID:99066935

A:Accession: T31113  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1832 <BAR>  
A:Cross-references: EMBL:AF068065; NID:g4063041; PID:g4063042; PIDN:AAC98153.1

Query Match 8.0%; Score 175.5; DB 2; Length 1832;  
Best Local Similarity 22.9%; Pred. No. 0.002;  
Matches 93; Conservative 45; Mismatches 176; Indels 93; Gaps 13;

Qy 43 TSANPGDFTFLNMPCEVFKYTTSTQTSVDLTADGVKATCQPSGEEFTFTSLTCTVNDAL 102  
Db 447 TKKPTT 506  
Qy 103 KSKIRAFCTVLPPIAFNVGGTSGSSDNL-----EDSKCFAGTNVTFNDGDKDI 151  
Db 507 PTTTITATTT-----TTSESVIKPDENCKWLEKNGCEAKGATYVGVIGKDGRI 557  
Qy 152 SIDVEFEKSTVDPSAYLIASRVMSLNVKVTTLFVAPOCENGYSGTMGFSSSNGDAI-- 209  
Db 558 ENGMATMIPNDTHVRFKVKDGNVISV-----RCGKG--AGKLEPPDRSLDTIPP 610  
Qy 210 -----DCSNIHIGITKGLNDWNPVSSSESVKTKTSNGIOIKY 249  
Db 611 VAGHNSCS-IIVGVGGGKIHVSPYSGKDVSLISAPIQCELENEVICDTCTA-----KY 664  
Qy 250 QNVAGYRPFIDAYISATNDYTNQVTLAYTNDYTCAGSLQSKPFTLRMTGKNSDAGSNGI 309  
Db 665 GAHSGYQTSADFVTTT-----AKPT-----TTTGAPGQ 695  
Qy 310 VIVATRTVOSTAVTTLFNPNSVDKTKTIEILOPIPTTTTTSYVGVVTSYLTKTAPI 369  
Db 696 PTTTGTGSPSKPTTTTT-----KATTTTILNPIITTT-TQKPTTTTTTKVPGKPP 747  
Qy 370 GETATVVDVVPYHTTTTVTSEWGTITTTTTTTTNTDSTIDTVVYVQVP 416  
Db 748 ATTTTLKPIVTTTTTKATTTTTTTPVTTTT-ATKRDEMTTTTTPLP 793

RESULT 8  
T40778  
hypothetical 129.5 kd protein - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
R:Accession: T40778  
R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Devlin, K.; Churcher, C.M.  
submitted to the EMBL Data Library, February 1998  
A:Reference number: Z21884  
A:Accession: T40778  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-973 <LYN>  
A:Cross-references: EMBL:AL021837; PIDN:CAAL7032.1; GSPDB:GN00067; SPDB:SPBC947.04  
A:Experimental source: strain 972h-; cosmid c947  
C:Genetics:  
A:Gene: SPDB:SPBC947.04  
A:Map position: 2

Query Match 8.0%; Score 174.5; DB 2; Length 973;  
Best Local Similarity 22.6%; Pred. No. 0.001;  
Matches 112; Conservative 63; Mismatches 157; Indels 163; Gaps 25;

Qy 15 TWSNANVAFKPGCYPTNNAVLGWSLDGTSANPGDFTFLNMPCEVFKYTTSTQTSV-DLTAD 73  
Db 149 TWSTNL-----PINPTTAI--YTSGSS-----NITPYSNRITNSVNDITSK 193  
Qy 74 --GVKATCQPSGEEF-----TTFSTLTCTV-----NDALASSIKAF 109  
Db 194 YLSVGTILTTISGSLDITSTFPANGTSGIVVEVITFAGTAVTAVSGSELYTSTFPAN 253  
Qy 110 GTV--TLPIAFNVGGTSGSTDLSDSKCF-----TAGTNTVTFNDGD 148

```
Db 254 GTTSCVEVVPAGTRVTKISGKFFTTTTTDDASGVTSQVEVVLPTAGTNTMTVVSQS 313
Qy 149 K-----DISIDVEFEKSTVDPSAYLYASR-----VMSPLN 178
Db 314 REFTSVVSGTSGEIVVPTAGVMVETIVSGSEIENFTYPASGTRTGVEVVIPTAG 373
Qy 179 KVT-----TLFVAPOCENGYSGTGMSNGDVAIDCSNIHIGITKGLNDWNPVSS 231
Db 374 TVTETISGSELYSTFPANGTSGV-----EVIPTAGTRT-VTK-----ISG 417
Qy 232 ESFSYTKTCTSGNIGIYQNPAGYRPFIDAYISATDVNQYTLAYTNDYT-CAGSRLOSK 290
Db 418 SKF-FTTTDASGT-----VSGTEVVLPCTAGTNTMTVVSQSRF--- 455
Qy 291 PFTLRWTKYKNSDAGSNG--IIVVATRTVTDSTTA-----VTTLPFNPSVDKTKTIEL 343
Db 456 -----FTSVVSASGTSGEIIIEPTAGVGTETVVSQVGYTTTTPAHDVTV--SGTVEVV 508
Qy 344 QP---IPTTTITTSYGVVTSYLTKTAPIGETATVIVDVPYHT---TTTTSWTGTT 396
Db 509 EPTAGVVTETVVSQVGYTTAY-----PAHDTVSGTEVVETVETVVSQVGYTT 563
Qy 397 TTTTTRTNPDSIDTV 411
Db 564 TYPADTVSGTEVV 578

RESULT 9
T38221
hypothetical serine-threonine rich protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 17-Nov-2000
R:Murphy, L.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z21780
A:Accession: T38221
A:Molecule type: DNA
A:Residues: 1-456 <MUR>
A:CROSS-references: EMBL:AL021813; PIDN:CAA16975.1; GSPDB:GN00066; SPDB:SPAC23A1.01c
A:Experimental source: strain 972h-; cosmid c23A1
C:Genetics:
A:Gene: SPDB:SPAC23A1.01c
A:Map position: 1
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase

Query Match 7.8%; Score 171.5; DB 2; Length 456;
Best Local Similarity 24.8%; Pred. No. 0.0006;
Matches 110; Conservative 57; Mismatches 147; Indels 129; Gaps 24;

Qy 10 SFNSLWSNAANYAFKPGYPTWNAVLGSLDGTSAHPGDTFTLNNPCVFKYTTQ----- 65
Db 94 SPTSIR-DDSVIYAQSGTFY-----IVGG--EGISSTTGTF-----QSMFTTSSQTSNG 141
Qy 66 -----TSVDLTADGVY--ATCOF-YSGEFTTFTLCTVNDAL-----KSSIK 107
Db 142 HASASTSIPSTAITVNSTIYSSATSSFPYS---TDVSVSTGTSIDVILPPASSFS 197
Qy 108 AFGVTPLPIAFNVGGTSGSDLEDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 167
Db 198 SFSITIT-----NTSMIPSSSSFTTTGSPYNTS-----SPLPSSVLISSAS 238
Qy 168 LYASRVMP-----SLNKVTTTLFVAPOCENGYSCTMGFSSNGDVAIDC 211
Db 239 LSSSVLPSTSIITSTPTVTVSSSSLSSTFPSTYNTLTGTTTGTGATVSSPPFYS 298
Qy 212 SNHIGITKGLNDWNPVSSSEFSYKTCISNGIQIYQNPAGYRPFIDAYISATDVNQ 271
Db 299 SVIPTSVPSSVSF-----TSSSSSTITTLTASNTSVYTTGTGTG-----SAT----- 341
Qy 272 YTLA----YTNDY---TCAGSRLOSKFTFLRWTKYKNSDAGSNGIIVATRTVTDSTTAV 325
```

```
Db 342 FTSPPPFYSSNVIPTSVPSVSS--FTSSNSSYTTTLTASN--TSITYTGTGTSATFT 397
Qy 326 TITLFF--NPSVDKTKTIELIQPIPTT--TITTSYGVVTSYLTKTAPIGETATVIVDVPY 381
Db 398 SSPPFYSNSV-----IPTSPSSVSSTSSNSYTT-----TLTA----- 433
Qy 382 HTTTTTSWTGCTGIITTTTRTP 404
Db 434 -SNTTTFRTGTGTSATFTSSPP 455

RESULT 10
T21389
hypothetical protein F26C11.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T21389
R:Matthews, P.
submitted to the EMBL Data Library, December 1994
A:Reference number: Z19416
A:Accession: T21389
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1251 <WIL>
A:CROSS-references: EMBL:Z47072; PIDN:CAA87369.1; GSPDB:GN00020; CESP:F26C11.3
A:Experimental source: clone F26C11
C:Genetics:
A:Gene: CESP:F26C11.3
A:Map position: 2
A:Introns: 24/1; 111/1; 208/1; 272/1; 380/1; 394/3; 485/3; 586/3; 630/3; 669/3; 713/3

Query Match 7.8%; Score 171; DB 2; Length 1251;
Best Local Similarity 23.8%; Pred. No. 0.0024;
Matches 107; Conservative 60; Mismatches 154; Indels 128; Gaps 23;

Qy 27 PGYPTWNAVLGSLDGTGTS-----ANPGDTFTLNNPCVFKYTTQSVDLT 71
Db 609 PGTTTN-----WPTGGTTRMLPSGEIILSESIAIYPNCT-TVLMQLIYTFSTNKTRETT 663
Qy 72 AD--GVKYATCOFYSGEFT-----TFSTLTCTVN-----DALKSSIKAF- 109
Db 664 TDTEGCKTKTSTISSSSKFSITPTPTPSGGTTYNMPTGGTTRTLPSGEIILSESIAFQ 723
Qy 110 --GTVTLPFAFNVGGTSGSDLEDSKCFAGTNTVTFNDGDKDISIDVEFEKSTVDPSAY 167
Db 724 NCTTVLMQLIYN-----PSTN-----KTRTETTTDAGCKKTSSTSKISTTPTSPT-- 769
Qy 168 LYASRVMPSLNKVTTTLFVAPOCENGYSCTMGFSSNGDVAIDCSNIHIGITKGLNDWNY 227
Db 770 --SSKPTPTSTSMITTYNWP---TGGTTRTL---PSGEI----- 800
Qy 228 PVSESEFSYTKCTSGNIGIYQNPAGYRPFIDAYISATDVN--OYTLAVTNDYTCAGS 285
Db 801 -ILSESIAIYKNCITVLMQLIYN--PSTNKTRE---TTTDAQGCKATSSTSLSKPTSPSS 854
Qy 286 RLQSKP-FTLRW--TCYKNSDAGSNGIIVATRTVTDSTTAVTTLPRNPVSDKTKT---- 339
Db 855 STASPTPTIYNPTGGTTRTLPSGEIILSESIAIYKNCITVLMQLIYNPSTNKTRETT 914
Qy 340 -----IEILOPIPTTITTSYGVVTSYLTKTAPIGE-----TATVIVDVPY 381
Db 915 DAQCKATITPTP---ITTYI--NWPQTGTRTLPSGEIILSESIAIYKNCITVLMQLIY 970
Qy 382 H-----TTTTVTSEWTGTTTTTTRTNP 405
Db 971 NPSTNKTRETTSSDAQCKATSTTQPTT 999

RESULT 11
A45155
mucin FM-C.1 - African clawed frog (fragment)
```

C:Species: Xenopus laevis (African clawed frog)  
C:Date: 26-May-1994 #sequence\_revision 26-May-1994 #text\_change 21-Jul-2000  
C:Accession: A45155  
R:Hauser, F.; Hoffmann, W.  
J. Biol. Chem. 267, 24620-24624, 1992  
A:Title: P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1)  
A:Reference number: A45155; MUID:93077556  
A:Accession: A45155  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-662 <HAU>  
A:Cross-references: GB:L02115; MID:g214147; PIDN:AAA74725.1; PID:g951460  
C:Superfamily: trefoll homology  
F:162-207/Domain: trefoll homology <TRF1>  
F:107-347/Domain: trefoll homology <TRF2>  
F:354-394/Domain: trefoll homology <TRF3>  
F:526-566/Domain: trefoll homology <TRF4>  
F:573-613/Domain: trefoll homology <TRF5>  
F:621-661/Domain: trefoll homology <TRF6>

Query Match 7.6%; Score 167.5; DB 2; Length 662;  
Best Local Similarity 24.3%; Pred. No. 0.0018;  
Matches 94; Conservative 28; Mismatches 166; Indels 99; Gaps 16;

QY 42 GTSANPGDTFLNMPGVFKYTSQTSVDLTADGVKATCFYSGEEFTTSTLTCTVND 101  
DB 206 GHSHEHTTTTAKPTGIIQTITPTTTTTRK-----ATPTTTTTRKATPT 254  
QY 102 LKSSIKAFGVTLPIAPNVGGTSGDLSKCFKTAGTNTVTFNDGDKDISIDVEFEKST 161  
DB 255 TTTTTRKATTTTTP-----TTTTTTTTRKATTTTPTT-----TT 286  
QY 162 VDPISAYLVASRWPSLNVKVTLLFVAPQCENGYSGTMGSSNGSDVAIDCSNIHIGTKG 221  
DB 287 TTPPT-----TTTKATTTT-----TSGECKMEPSKRE---DCG--YSGITES 325  
QY 222 LNDMNPVSSSESYTKTCTSGNIGI-----KYQNPAGYRPFIDAYISATDVNQTILA 275  
DB 326 QCRTKGCCFSSIPQTKWCFYLSQVADCKVEPSQRVDCGFRG-----ITADQCRKNCC 380  
QY 276 YTNDYTCAGSLQSKPFTLVGYNKSDAGSNGIVIVATRTVTDSTAVTTLPFNPSVD 335  
DB 381 F--DSSISGT-----KWCFTSTQ-----VAATKTTTPTT--TTTPTTTTTT 419  
QY 336 K----TKTIELQPIPTTTTTSYGVVTSYLTAKTAPIGETATVIVDVPYHTTTTSTSEW 391  
DB 420 KATTTTPTTTTPTTTTPTTTT--TKATTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTT 478  
QY 392 TGTITT--TTTRTNPTDSIDTVVQVP 416  
DB 479 TKATTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTT 505

RESULT 12  
S48992  
floculation protein homolog YHR211w - yeast (Saccharomyces cerevisiae)  
C:Species: Saccharomyces cerevisiae  
C:Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 29-Oct-1999  
C:Accession: S48992  
R:Naciri, C.  
submitted to the EMBL Data Library, February 1994  
A:Description: The sequence of S. cerevisiae cosmid 9177.  
A:Reference number: S46671  
A:Accession: S48992  
A:Molecule type: DNA  
A:Residues: 1-1075 <MAC>  
A:Cross-references: EMBL:U00029; NID:g551322; PID:g458919; GSPDB:GN00008; MIPS:YHR211w  
C:Genetics  
A:Gene: MIPS:YHR211w  
A:Map position: 8R

Query Match 7.4%; Score 161.5; DB 2; Length 1075;  
Best Local Similarity 22.2%; Pred. No. 0.0083;  
Matches 110; Conservative 57; Mismatches 209; Indels 119; Gaps 22;

QY 21 NFAKPGPG-YPTWNAVIGNSLD--GTSANPGDTFLNAN-----PCVFKYTTTSQTSVDLT 71  
DB 104 NMGCKGACGMSNSQGIAYNSTDLFCGYTPTNV-TLEMTGYFLPQPGSYTFSFATVDD 162  
QY 72 A-----DGVKATCFYSGEEFTTSTLTCTVN-----  
DB 163 AILSVGGSIAFECC---AOEQPITSTNFTINGIKWDGSLPDNITGVVYAGYVYPL 218  
QY 100 DALKSIIKAFCTVLPFAFNVGCTGSSV-----DLEDSKC-----FTAGNTV 142  
DB 219 KVVSNAYNSWGLTFLISVELPDGTVVSDNFEYVYSPDDLSQSNCTIPDPSIHTTSTITT 278  
QY 143 -----TFNDGDKDISIDVEFEKSTVDPISAYLV-----ASRWPSLNVKVTLLFVAPQCE 190  
DB 279 TTEPWTGFTSTSTEMTITDNGQLTDETVIVIRPTTASITTTTTEPWGTGTSTSTE 338  
QY 191 NGYTSQTMGFSNGSDVAIDCSNIHIGTKGLNDNPNVSSSEFSYTKTCTSGNIGIYKQ 250  
DB 339 MTTVTGTNGQPTDETVIVIRPTSEGLITTTTPTTGTCTSTSTEMTIVTGTNG--QPTDE 397  
QY 251 NVPAGYRPFIDAYISATDVNQTILAYTNDYTCAGSLQSKPFTLRWTGYNKSD----- 303  
DB 398 TVIVRTPTSEGLITTT-----TEPWGTGTSTSTEV-----TITGNGQPIDETVIVIR 448  
QY 304 -AGSNGIVIVAT---TFTVTDSTAVTLPFNPSVDKTKTIELQP-----IPTT----- 349  
DB 449 TPTSEGLITTTTTEPWGTGTSTSTEMTIVTGTNGQPTDETVIVIRPTSEGLITTTTTEPW 508  
QY 350 -TITTSYGVVTSYLTAKTAPIGETATVI--VDVPYHTTTTSTSEWGTGIT-----TTTT 400  
DB 509 TGTFTSTSTEVTTTGTNGQPTDETVIVIRPTSEGLITTTTTEPWGTGTSTSTEMTIVT 568  
QY 401 RTN--PTDSIDTVV 413  
DB 569 GTNGQPTD--ETVIV 581

RESULT 13  
AC2224  
hypothetical protein all3346 [imported] - Anabaena sp. (strain PCC 7120)  
C:Species: Anabaena sp.  
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
C:Accession: AC2224  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2224  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1428 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA075045.1; PID:g17132441; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics  
A:Gene: all3346

Query Match 7.4%; Score 161; DB 2; Length 1428;  
Best Local Similarity 24.5%; Pred. No. 0.013;  
Matches 118; Conservative 43; Mismatches 191; Indels 130; Gaps 24;

QY 13 SUTWSNAN--YAFKPGPYTNNAVIGNSLQTSANPGD-----TFTLNMP 56  
DB 808 SVTEGDTNLIYTFRTGTSTTNALVYNSVAGTATLNTDYAQTGAASFTATTGTITFVAG 867  
QY 57 CVFKYTTTSQTSVDLTADGVKATCFYSGEEFTTSTLTCT-----VNDALKS-----S 105

Db 868 ASTALLINPTADTTVESNETVALTASGTYGTVGTTAVTGTITINDPFSITLAVSPAS 927  
QY 106 IKACETV-----TLPANVGCTGS-STDLDS--KCFAGTNVTVNDGD 148  
Db 928 VTEDGTPNLYITRTGTTALTINFGAGTALINDYAAQASAFRTTGTITFAA 987  
QY 149 KDISIDVEFEKSTVDSAYLYASRWPSLKNKVTFLFVAPOCENGTSMTGFSSSNGDVA 208  
Db 988 STAIL-----TINPTA-----DTTVESNETVALTLA--SGTGYVGT--TAVGTIT 1031  
QY 209 ICSNIHIGITGLDNMYPV-----SSESFSYTKTCTSNQIQIKYQNVAPAGYRPFIDAYI 264  
Db 1032 ND-DTLPGTITINLGSOTIVEGNSPQNVYTVTLQASQI-----ITVQYA 1079  
QY 265 SATDNOYTLAYTNDYTCAGSRLOSKP-----FTLRWGYKNS 302  
Db 1080 TANG-----TATAGSDYTSTGTLTFNPGETSKVINIPILNDSVNEANETFLRTSPTNA 1135  
QY 303 DAGSNGIVIVATRVTD--STTATVTLFPNPSVDKTKTIEILOPIPTTIT--TSYVGVT 359  
Db 1136 TLGTNTV-----TTTITDLSASVTTTLPTN-----VENLTGTAAINGTGNAG-- 1181  
QY 360 TSYLKTAPIGETATVVD--VPYHTTTTTSVSEMTGTTTITTTTNRNPTDSID----TVVV 413  
Db 1182 NNILT-----GNSGNILSGGAGNDYAFVANAALGTDITETATGGIDTIDFNGSTATV 1236  
QY 414 QV 415  
Db 1237 RV 1238  
RESULT 14  
T22808  
hypothetical protein F56H9.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22808  
R:Burton, J.  
submitted to the EMBL Data Library, June 1996  
A:Reference number: Z19618  
A:Accession: T22808  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-770 <WIL>  
A:Cross-references: EMBL:Z74473; PIDN:CAA98949.1; GSPDB:GN00023; CESP:F56H9.1  
A:Experimental source: clone F56H9  
C:Genetics:  
A:Gene: CESP:F56H9.1  
A:Map position: 5  
A:Introns: 235/1; 262/2; 320/1; 367/2; 510/3; 654/1; 681/2  
Query Match 7.1%; Score 156.5; DB 2; Length 770;  
Best Local Similarity 20.3%; Pred. NO. 0.011;  
Matches 106; Conservative 48; Mismatches 170; Indels 197; Gaps 17;  
QY 50 TFLNMPCKVFKYTTQSVDLTADGVKYATCFYSGEEFTTFLTCTVNDALSKSKAF 109  
Db 166 TTTTTPVATTSTATTVPPTTSTTTT-----PPTTSTTTTTPPTTSTTTT 220  
QY 110 GTVTL-----IAPNVGTT-----GSDTL 129  
Db 221 TTTTTLPTTTTMMIPCSLDSLTGKDNFPQVDIDVTYSNLVANIIPGMDMQSLSTM 280  
QY 130 EDSKCFAGTN-----TVTFNDGDKDISID-----VEPEKST-- 161  
Db 281 QISCTAIAAGNAYMIFDQVPAENQOFPQSVTVNAOCNRIINLPMFEKLSYKTSFV 340  
QY 162 -----VDPSAYLYASRWPSLN-----KVTLTFV- 185  
Db 341 ILCTSKLSNFILLTVYFLYIGPTELLYTGISNMNLNTYLNQTMGLWLOGPITQMILT 400  
QY 186 -----APOCENGYSCTM-----GFSSNG-DVAIDCSNI 214

Db 401 INRELVIWTPPHVPOISHRITLIGALSWSWITVTLSTLIGLPGDNNNFIDIIINLANC 460  
QY 215 HIGITKGLDNMYPVSSSEFSYTKTCTSNQI-----KYQNVAPAGYRPF-- 259  
Db 461 RV-----PIGFEHIGYSPCNQKITIIVSGIFLGFLTNFMFMIGGLIYT 509  
QY 260 -----TDAVIS-----ATDNOYTLAYTNDY--TCAGSRLOSKPFTLRWGYKNSD 303  
Db 510 WKTLTIFRLSNFKCVMKWSTLLHLLTYKLDVINSCIATSPADPTT-----TTT 562  
QY 304 AGSNGIVIVATRVTDSTAVTTLFPNPSVDKTKTIEILOPIPTTITTSYVGVTSYL 363  
Db 563 EATTTTEITTTTTEEVTTTTPPTTSTTTTSTTTTSTTTTSTTTTSTTTTSTTTT 622  
QY 364 TKTAPIGETATVIVDVPYHTTTTTSVSEMTGTTTITTRNP 404  
Db 623 TTTTPTTTST-----TTTTTTTTTATPTTTTTTTPP 655  
RESULT 15  
T22696  
hypothetical protein F55B11.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T22696  
R:Alnscough, R.  
submitted to the EMBL Data Library, December 1996  
A:Reference number: Z19601  
A:Accession: T22696  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-851 <WIL>  
A:Cross-references: EMBL:Z83318; PIDN:CA805903.1; GSPDB:GN00022; CESP:F55B11.3  
A:Experimental source: clone F55B11  
C:Genetics:  
A:Gene: CESP:F55B11.3  
A:Map position: 4  
A:Introns: 49/3; 123/3; 226/1; 282/3; 669/3; 743/3  
Query Match 7.1%; Score 156; DB 2; Length 851;  
Best Local Similarity 23.0%; Pred. No. 0.014; Indels 90; Gaps 11;  
Matches 87; Conservative 32; Mismatches 170; Indels 90; Gaps 11;  
QY 43 TSANPGDPTFLNMPCKVFKYTTQSVDLTADGVKYATCFYSGEEFTTFLTCTVNDAL 102  
Db 330 TTPSELSTTASVP-----TTTTSVPTTPTTTPPTT-----TTVPTTPTTSTTT 373  
QY 103 KSSIKAFGVTLPLIAFNVGTSSTDLDSKCFCTAGTNVTVNDGDKDISIDVEFEKSTV 162  
Db 374 -----TTTVPT-----TTTTPPTTPTTPTTPTTPTT-----VPTTPTTV 405  
QY 163 DPSAYLYASRWPSLKNKVTFLFVAPOCENGYSMTGFSSSNGDVAIDCSNIHIGITKGL 222  
Db 406 P-----PTTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 440  
QY 223 NDNMYPVSSSEFSYTKT-----CTSNQIQIKYQNVAPAGYRPFIDAYISATDNOYTLAYTND 279  
Db 441 -----PTTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 485  
QY 280 YTCAGSRLOSKPFTLRWGYKNSDAGSNGIVIVATRVTDSTTAVTTLFPNPSVDKTKT 339  
Db 486 TTTVPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 545  
QY 340 IEILOPIPTTPTTTSYVGVTSYLTKTAPIGETA--TVIVDVPYHTTTTTSVSEMTGTTT 397  
Db 546 PTTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 604  
QY 398 TTRTPTDTSIDTVVQVP 416  
Db 605 TTTVPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 623

Fri Oct 4 08:13:54 2002

us-09-715-876-8\_copy\_17\_432.text.rpr

Page 8

Search completed: October 3, 2002, 16:30:32  
Job time: 135 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:30:07 ; Search time 13.4 Seconds  
(without alignments)  
1202.040 Million cell updates/sec

Title: US-09-715-876-8\_COPY\_17\_432

Perfect score: 2190

Sequence: 1 AKTIGVDFSNLSLWSNAA.....TTTTTNPNTSDIDVVVQVP 416

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2178	99.5	1260	1	ALSI1_CANAL
2	1907	87.1	1419	1	ALAI1_CANAL
3	1889	86.3	1119	1	ALSI3_CANAL
4	1773.5	81.0	468	1	ALSI2_CANAL
5	1600	73.1	469	1	ALSI4_CANAL
6	308.5	14.1	650	1	SAG1_YEAST
7	177	8.1	1322	1	YAG3_YEAST
8	177	8.1	1537	1	FLQ1_YEAST
9	171	7.8	1251	1	YQJ3_CAEEL
10	167.5	7.6	662	1	MUC1_XENLA
11	161.5	7.4	1075	1	FLQ5_YEAST
12	160	7.3	567	1	CHI3_CANAL
13	156	7.1	540	1	GUX1_ASPAC
14	154.5	7.1	1367	1	AMVH_YEAST
15	153.5	7.0	725	1	AGAL_YEAST
16	152	6.9	1902	1	P1P_LACLC
17	148.5	6.8	1169	1	YK82_YEAST
18	145	6.6	995	1	YI09_YEAST
19	145	6.6	1041	1	EGT2_YEAST
20	145	6.6	1609	1	FIQ2_YEAST
21	145	6.6	1902	1	P3P_LACLC
22	144.5	6.6	881	1	YJH8_YEAST
23	143	6.5	1902	1	P2P_LACLC
24	142	6.5	522	1	NU62_HUMAN
25	141	6.4	1902	1	P2P_LACPA
26	138	6.3	605	1	YK68_YEAST
27	135.5	6.2	400	1	MU1_XENLA
28	135	6.2	670	1	VG50_HSVII
29	133	6.1	827	1	CSG_HALVO
30	132.5	6.1	562	1	CHIT_YEAST
31	132.5	6.1	796	1	Y58A_CAEEL
32	131.5	6.1	1306	1	MSB2_YEAST
33	130.5	6.0	338	1	SED1_YEAST

34	130	5.9	535	1	NU62_RAT
35	128	5.8	907	1	VGP3_EBV
36	128	5.8	5179	1	MUC2_HUMAN
37	127.5	5.8	598	1	YAU1_SCHPO
38	127.5	5.8	886	1	VGP3_EBVA8
39	127	5.8	600	1	SP96_DICDI
40	126.5	5.8	1802	1	HKR1_YEAST
41	126	5.8	445	1	WAPA_STRMU
42	125	5.7	767	1	AMVH_SACDI
43	125	5.7	1140	1	YM96_YEAST
44	125	5.7	1161	1	DAN4_YEAST
45	124.5	5.7	768	1	AMV1_SACDI

ALIGNMENTS

RESULT 1	ALSI1_CANAL	STANDARD;	PRT; 1260 AA.
ID	P46590;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Agglutinin-like protein 1 precursor.		
GN	ALSI		
OS	Candida albicans (Yeast).		
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.		
OX	NCBI_TaxID=5476;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=ATCC 11651 / B792;		
RX	MEOLINE=95272392; Pubmed=7752895;		
RA	Hoyer L.B., Scherer S., Shatzman A.R., Livi G.P.;		
RT	"Candida albicans ALSI: domains related to a Saccharomyces cerevisiae		
RT	sexual agglutinin separated by a repeating motif.";		
RL	Mol. Microbiol. 15:39-54(1995).		
CC	-1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.		
CC	-1- PFM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).		
CC	-1- SIMILARITY: TO YEAST SAG1.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).		
CC	-----		
CC	EMBL; L25902; AAC41649.2; -		
DR	Cell adhesion; Glycoprotein;		
KW	POTENTIAL.		
FT	SIGNAL		
FT	1		
FT	CHAIN		
FT	18 1260		
FT	DOMAIN		
FT	433 792		
FT	REPEAT		
FT	433 468		
FT	REPEAT		
FT	459 504		
FT	REPEAT		
FT	505 540		
FT	REPEAT		
FT	541 576		
FT	REPEAT		
FT	577 612		
FT	REPEAT		
FT	613 648		
FT	REPEAT		
FT	649 684		
FT	REPEAT		
FT	695 720		
FT	REPEAT		
FT	731 756		
FT	REPEAT		
FT	757 792		
FT	DOMAIN		
FT	983 1152		
FT	REPEAT		
FT	983 1043		
FT	REPEAT		
FT	1092 1152		
FT	DOMAIN		
FT	399 404		
FT	DOMAIN		
FT	408 418		
FT	DOMAIN		
FT	450 455		
FT	DOMAIN		
FT	486 491		
FT	DOMAIN		
FT	522		



ALS3 CANAL STANDARD; PRT: 1119 AA.  
AC 07423; 1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Agglutinin-like protein 3 precursor.  
GN ALS3.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=3476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RA MEDLINE=98309840; PubMed=9644209;  
RX Hoyer L.L., Payne T.L., Bell M., Myers A.M., Scherer S.;  
RT "Candida albicans ALS3 and Insights into the nature of the ALS gene family.";  
RL Curr. Genet. 33:451-459(1998).  
CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; 087956; AAC39486.1; --  
KW Cell adhesion; Glycoprotein; Repeat; Signal.  
DR SIGNAL 1 17  
FT CHAIN 18 1119 AGGLUTININ-LIKE PROTEIN 3.  
FT DOMAIN 433 792 10 X 36 AA TANDEM REPEATS.  
FT REPEAT 433 468 1-1.  
FT REPEAT 469 504 1-2.  
FT REPEAT 505 540 1-3.  
FT REPEAT 541 576 1-4.  
FT REPEAT 577 612 1-5.  
FT REPEAT 613 648 1-6.  
FT REPEAT 649 684 1-7.  
FT REPEAT 685 720 1-8.  
FT REPEAT 721 756 1-9.  
FT REPEAT 757 792 1-10.  
FT DOMAIN 399 404 POLY-THR.  
FT DOMAIN 450 455 POLY-THR.  
FT DOMAIN 557 563 POLY-THR.  
FT DOMAIN 593 597 POLY-THR.  
FT DOMAIN 630 635 POLY-THR.  
FT DOMAIN 666 671 POLY-THR.  
FT DOMAIN 702 707 POLY-THR.  
FT DOMAIN 738 743 POLY-THR.  
FT DOMAIN 774 777 POLY-THR.  
FT DOMAIN 1044 1047 POLY-THR.  
FT CARBOHYD 471 471 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 543 543 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 579 579 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 651 651 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 687 687 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 723 723 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 845 845 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 987 987 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 1061 1061 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 1119 AA; 119927 MW; 6A3FB3FC8C879A71 CRC64;

Query Match 86.3%; Score 1889; DB 1; Length 1119;  
Best Local Similarity 84.9%; Pred. No. 1.6e-120;  
Matches 353; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

QY 1 AKTITGVDFSPNSLTWNSAANAAYAFKPGYPTWNAVLGNSLDGTSANPGDFTTLNMPGVFK 60  
DB 17 KTTITGVDFSPNSLTWNSAANAAYAFKPGYPTWNAVLGNSLDGTSASPGDFTTLNMPGVFK 76  
QY 61 YTSQTSVDLTADGVKYATCOFYSGEFTTCTVNDALKSIRAKFATVTLPIAFNV 120  
DB 77 FTSQTSVDLTADGVKYATCOFYSGEFTTCTVNDALKSIRAKFATVTLPIAFNV 136  
QY 121 GGTGSSVDLEDSKCTAGTNTVTFNDGDKDISIDVEFEKSTVDFPSAYLYASRVMPSLNKV 180  
DB 137 GGTGSSVDLEDSKCTAGTNTVTFNDGDKDISIDVEFEKSTVDFPSAYLYASRVMPSLNKV 196  
QY 181 TTFVAPQCEGYTSGTMGFSSNGDVAIDCSNIHIGITKGLNDWNPVSESEFSYKTC 240  
DB 197 STLFVAPQCEGYTSGTMGFSSNGDVAIDCSNIHIGITKGLNDWNPVSESEFSYKTC 256  
QY 241 TSNIGIQIKYQNPAGYRPFVDAYISATDVNQYTLAYTNDYTCAGSRQSKPFTLRWTGYK 300  
DB 257 SSNGIFITKYNPAGYRPFVDAYISATDVNQYTLAYTNDYTCAGSRQSKPFTLRWTGYK 316  
QY 301 NSDAGSNGIVIVATRTVTDSTTAVTTLPPNPVSDTKTIEILOPIPTTTTTSYGVVTT 360  
DB 317 NSDAGSNGIVIVATRTVTDSTTAVTTLPPNPVSDTKTIEILOPIPTTTTTSYGVVTT 376  
QY 361 SYLTKTAPIGETATVVDVYPHTTTTSEWTTTITTTTTRTNPDTSDITVYVQVP 416  
DB 377 SYLTKTAPIGETATVVDVYPHTTTTSEWTTTITTTTTRTNPDTSDITVYVQVP 432  
RESULT 4  
ALS2 CANAL ID ALS2 CANAL STANDARD; PRT: 468 AA.  
AC 074657;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Agglutinin-like protein 2 precursor (Fragment).  
GN ALS2.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RX MEDLINE=98440424; PubMed=9765564;  
RA Hoyer L.L., Payne T.L., Hecht J.E.;  
RT "Identification of Candida albicans ALS2 and ALS4 and localization of  
RT als proteins to the fungal cell surface.";  
RL J. Bacteriol. 180:5334-5343(1998).  
CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AF024580; AAC64235.1; --  
KW Cell adhesion; Glycoprotein; Repeat; Signal.  
DR SIGNAL 1 17  
FT CHAIN 18 468 AGGLUTININ-LIKE PROTEIN 2.  
FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 315 315 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT NONTER 468 468  
SQ SEQUENCE 468 AA; 50203 MW; BFE773E169ED0FAF CRC64;

Query Match 81.0%; Score 1773.5; DB 1; Length 468;  
Best Local Similarity 80.3%; Pred. No. 3.6e-113;

Matches 334; Conservative 27; Mismatches 54; Indels 1; Gaps 1;

QY 1 AKTITGVDFDSNSLTWSNAANYAFKPGYPTWNAVLGWSLDGTSANPGDFTFLNMPGVFK 60  
DB 17 AKVITGVDFNSLTWTRAGNAYKGNRPWNAVLGWSLDGTSANPGDFTFLNMPGVFK 76  
QY 61 YTTSTQTSVDLTADGVKATQFYSGEERTFTSTLTCTVDALKSSIKAFKGTVTLPIAFNV 120  
DB 77 FITDQTSVDLTADGVKATQFYSGEERTFTSTLTCTVDALKSSIKAFKGTVTLPIAFNV 136  
QY 121 GGTGSGTDLKSCFTAGTNTVTTFNDGDKDISDVEFEKSTVDPSAYLYASRVMPSLNKV 180  
DB 137 GGTGSGVDLANSQCFAGINTVTFNDGDKDISDVEFEKSTVDPSAYLYASRVMPSLNKV 196  
QY 181 TFLFVAPOCENGYSCTMGFSSNGDVAIDCSNHIHIGITKGLNDWNPVSSSESYTKTC 240  
DB 197 STIVFAPOCANGYSCTMGFSSNGDVAIDCSNHIHIGITKGLNDWNPVSSSESYTKTC 256  
QY 241 TNGTQIKYQNVAGYRPFIDAYISATDVNOYTLATNDYTCAGSLQSKRPFLLRWGTYK 300  
DB 257 STGISITVENFAGYRPFIDAYISATDVNOYTLATNDYTCAGSLQSKRPFLLRWGTYK 315  
QY 301 NSDAGSNGIVATRTVTDSTTAVTLPFNPSVDKTKTIEILQPIPTTTTTSYGVGVT 360  
DB 316 NSEANSNGFVATRTVTDSTTAVTLPFNPSVDKTKTIEILQPIPTTTTTSYGVGVT 375  
QY 361 SYLTKTAPIGETATVIVDVPYHTTTTSEMTGTTTTTTRNPTDSDITVVVQVP 416  
DB 376 SYSTKTAPIGETATVIVDVPYHTTTTSEMTGTTTTTTRNPTDSDITVVVQVP 431

RESULT 5  
ID ALS4\_CANAL STANDARD; PRT; 469 AA.  
AC 074660;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Agglutinin-like protein 4 precursor (fragment).  
GN ALS4.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98440424; PubMed=97655564;  
RA Hoyer L.L., Payne T.L., Hecht J.E.;  
RT "Identification of Candida albicans ALS2 and ALS4 and localization of  
RT als proteins to the fungal cell surface."  
RL J. Bacteriol. 180:5334-5343(1998).  
CC -!- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.  
CC -!- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF024584; AAC64239.1; .  
DR Cell adhesion; Glycoprotein; Repeat; Signal.  
KW SIGNAL  
FT CHAIN 1 17  
FT NON\_TER 18 >469 AGGLUTININ-LIKE PROTEIN 4.  
FT SEQUENCE 469 AA; 49604 MW; 0EDCAB19B89EFCB1 CRC64;  
SQ

Query Match 73.1%; Score 1600; DB 1; Length 469;  
Best Local Similarity 71.2%; Pred. No. 26-101;

Matches 296; Conservative 42; Mismatches 78; Indels 0; Gaps 0;

QY 1 AKTITGVDFDSNSLTWSNAANYAFKPGYPTWNAVLGWSLDGTSANPGDFTFLNMPGVFK 60  
DB 17 AKVITGVDFNSLTWNAASAYKGPATPTWNAVLGWSLDGATASAGDFTFLDMPGVFK 76  
QY 61 YTTSTQTSVDLTADGVKATQFYSGEERTFTSTLTCTVDALKSSIKAFKGTVTLPIAFNV 120  
DB 77 FITDQTSVDLTADGVKATQFYSGEERTFTSTLTCTVDALKSSIKAFKGTVTLPIAFNV 136  
QY 121 GGTGSGTDLKSCFTAGTNTVTTFNDGDKDISDVEFEKSTVDPSAYLYASRVMPSLNKV 180  
DB 137 GGTGSGVDLANSQCFAGINTVTFNDGDKDISDVEFEKSTVDPSAYLYASRVMPSLNKV 196  
QY 181 TFLFVAPOCENGYSCTMGFSSNGDVAIDCSNHIHIGITKGLNDWNPVSSSESYTKTC 240  
DB 197 VNLFLPQCAANGYSCTMGFSSNGDVAIDCSNHIHIGITKGLNDWNPVSSSESYTKTC 256  
QY 241 TNGTQIKYQNVAGYRPFIDAYISATDVNOYTLATNDYTCAGSLQSKRPFLLRWGTYK 300  
DB 257 TSTSVLVTQNVAGYRPFIDAYISATDVNOYTLATNDYTCAGSLQSKRPFLLRWGTYK 316  
QY 301 NSDAGSNGIVATRTVTDSTTAVTLPFNPSVDKTKTIEILQPIPTTTTTSYGVGVT 360  
DB 317 NSQAGSNGITVATRTVTDSTTAVTLPFNPSVDKTKTIEILQPIPTTTTTSYGVGVT 376  
QY 361 SYLTKTAPIGETATVIVDVPYHTTTTSEMTGTTTTTTRNPTDSDITVVVQVP 416  
DB 377 SYSTKTAPIGETATVIVDVPYHTTTTSEMTGTTTTTTRNPTDSDITVVVQVP 432

RESULT 6  
ID SAG1\_YEAST STANDARD; PRT; 650 AA.  
AC P20840;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alpha-agglutinin precursor (AG-alpha-1).  
GN SAG1 OR AGAL1 OR YJR004C OR J1418.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetales; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90014768; PubMed=2677666;  
RA Lipke P.N., Wojciechowski D., Kurjan J.;  
RT "AG alpha 1 is the structural gene for the Saccharomyces cerevisiae  
RT alpha-agglutinin, a cell surface glycoprotein involved in cell-cell  
RT interactions during mating."  
RL Mol. Cell. Biol. 9:3155-3165(1989).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288C;  
RX MEDLINE=90005993; PubMed=2676603;  
RA Hauser K., Tanner W.;  
RT "Purification of the inducible alpha-agglutinin of S. cerevisiae and  
RT molecular cloning of the gene."  
RL FEBS Lett. 255:290-294(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX STRAIN=S288C / FY1679;  
RA de Haan M., Smits P.H.M., Grivell L.A.;  
RN Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP PARTIAL SEQUENCE; CARBOHYDRATE-LINKAGE SITES, AND DISULFIDE BONDS.  
RX MEDLINE=96064684; PubMed=7592821;  
RA Chen M.-H., Shen Z.-M., Bobin S., Kahn P.C., Lipke P.N.;  
RT "Structure of Saccharomyces cerevisiae alpha-agglutinin. Evidence for  
RT a yeast cell wall protein with multiple immunoglobulin-like domains  
RT with atypical disulfides."  
RL J. Biol. Chem. 270:26168-26177(1995);



```

DR EMBL; U12980; AAC04971.1; -.
DR SGD; S0000059; YAL063C.
DR InterPro; IPR001389; Flocculin.
DR Pfam; PF00624; Flocculin; 13.
KW Hypothetical protein; Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 24
FT CHAIN 25 1322
FT TRANSMEM 366 388
FT TRANSMEM 754 775
FT CARBOHYD 135 135
FT CARBOHYD 187 187
FT CARBOHYD 203 203
FT CARBOHYD 257 257
FT CARBOHYD 262 262
FT CARBOHYD 270 270
FT CARBOHYD 329 329
FT CARBOHYD 419 419
FT CARBOHYD 464 464
FT CARBOHYD 509 509
FT CARBOHYD 554 554
FT CARBOHYD 599 599
FT CARBOHYD 644 644
FT CARBOHYD 689 689
FT CARBOHYD 734 734
FT CARBOHYD 886 886
FT SEQUENCE 1322 AA; 138072 MW; AADFDLFF13267CEA CRC64;

Query Match 8.1%; Score 177; DB 1; Length 1322;
Best Local Similarity 23.3%; Pred. No. 0.00017;
Matches 112; Conservative 53; Mismatches 181; Indels 134; Gaps 26;

QY 21 NYAFKGGPGYFNAVLG-NSLD--GTSANPGDFTFLNM-----PCVFYKTTTSQFS 67
DB 104 NGCKGGIGACSNPIIAYXWDLGFGYTPTNV-TLEMTGYLPQTGSGYTFKFAVDSDS 162
QY 68 VDLTADG-VKATCFQYSGEFTFTSLTCTVNDALKSSIKAF-----GTV----- 112
DB 163 AILVSGSIAFECC---AQEQPITSTNFIN-----GIKPNWSPDNIITGVVYAG 213
QY 113 -----TLPIAFNV-GGTGSGTDLSDSKFTAGTNTVFNDGKDISDV 155
DB 214 FYYPKVIYNAVAVGTLPISTLPDGTVDSDDFEG-----YVYTF-----DNLS- 259
QY 156 EPEKSTVDPSAIIYASRVMPSLNKVTLFVAPQENGYSGTGMFSSNGDVAIDCSNIH 215
DB 260 QPNCTIPDPSNYT-VSTTTITTEPWGTGFTSTSTEMTTVTGTNGVPTDETIVIRPTTA 318
QY 216 IGITKGLNDNYPVSEFSFYKTKTSNGIQKYQNVAGYRPFIDAYISATDQNYTLA 275
DB 319 STIITTEPNSTFTSTSTELTFTVTGTNGVTRDTETII-----VIRPTTATTAITTEP 372
QY 276 YTNDDYTCAGSRLOSKPFLRWYGYKNSDAGSNGI---VIVATRTVTDSTAVTL-PEN 331
DB 373 WNSTFTSTSTEL-----TTVTGNGLPDDETIIVIRPTTATTAITTAITTPQWN 419
QY 332 PSVDKTKT-----IEILOPIPTTITTT-----SYGVVTSYLTAKTA 367
DB 420 DFTSTSTELTFTVTGNGLPDDETIIVIRPTTATTAITTAITTAITTAITFTSTSTELTFTVG 479
QY 368 -----PIGETATVIVDVPYHTTT--TVTSEWGTGIT-----TTTTTRN--PTSDITVVV 413
DB 480 TNGLPDDET-IIIVIRPTTATTAITTAITTAITTAITTAITTAITTAITTAITTAITTAIT 536

RESULT
FL01_YEAST
ID FL01_YEAST STANDARD; PRT; 1537 AA.
AC P32768;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Flocculation protein FL01 precursor (Flocculin 1).

```

```

GN OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94262325; PubMed=8203162;
RA Watarai J., Takata Y., Ogawa M., Sahara H., Koshino S., Onnela M.-L.,
RA Alakainen U., Jaatinen R., Penttilae M., Keranen S.;
RT "Molecular cloning and analysis of the yeast flocculation gene FL01.";
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C / AB972;
RA MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RT "The nucleotide sequence of chromosome I from Saccharomyces
cerevisiae.";
RN [3]
RP PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=93289821; PubMed=8511970;
RA Teunissen A.W.R.H., Holub E., van der Hucht J., van den Berg J.A.,
RX Steensma H.Y.;
RT "Sequence of the open reading frame of the FL01 gene from
Saccharomyces cerevisiae.";
RN [4]
RP REVIEW.
RX MEDLINE=96076625; PubMed=7502576;
RA Teunissen A.W., Steensma H.Y.;
RT "Review: the dominant flocculation genes of Saccharomyces cerevisiae
constitute a new subtelomeric gene family.";
RN [5]
RP Yeast 11:1001-1013(1995).
CC -!- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
(Potential).
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X78160; CAA55024.1; -.
DR EMBL; L28920; AAC09499.1; ALT_SEQ.
DR PIR; S31230; S31230.
DR SGD; S0000084; FLO1.
DR InterPro; IPR001389; Flocculin.
DR Pfam; PF00624; Flocculin; 18.
KW Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
FT SIGNAL 1 24
FT CHAIN 25 1537
FT PROPEP 274 1240
FT DOMAIN 278 1087
FT DOMAIN 278 1087
FT REPEAT 323 367
FT REPEAT 368 412
FT REPEAT 413 457
FT REPEAT 458 502
FT REPEAT 503 547
FT REPEAT 548 592
FT REPEAT 593 637
FT REPEAT 638 682
FT REPEAT 683 727

```

FT REPEAT 728 772 1-11.  
FT REPEAT 773 817 1-12.  
FT REPEAT 818 862 1-13.  
FT REPEAT 863 907 1-14.  
FT REPEAT 908 952 1-15.  
FT REPEAT 953 997 1-16.  
FT REPEAT 998 1042 1-17.  
FT REPEAT 1043 1087 1-18.  
FT DOMAIN 1118 1157 2 X 20 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 1118 1137 2-1.  
FT REPEAT 1138 1157 2-2.  
FT DOMAIN 1226 1392 3 X 51 AA APPROXIMATE REPEATS.  
FT REPEAT 1226 1276 3-1.  
FT REPEAT 1291 1341 3-2.  
FT REPEAT 1342 1392 3-3.  
FT DOMAIN 1408 1434 3 X 9 AA APPROXIMATE TANDEM REPEATS.  
FT REPEAT 1408 1416 4-1.  
FT REPEAT 1417 1425 4-2.  
FT REPEAT 1426 1434 4-3.  
FT CARBOHYD 135 135 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 509 509 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 554 554 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 599 599 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 644 644 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 689 689 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 734 734 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 1114 1114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1537 AA; 160722 MW; 992650C6BE9A8CEA CRC64;

Query Match 8.18; Score 177; DB 1; Length 1537;  
Best Local Similarity 23.08; Pred. No. 0.00021;  
Matches 112; Conservative 47; Mismatches 181; Indels 146; Gaps 25;  
QY 21 NYAFKGPQ-YPTWNAVGLMSLD--GTSANPGDFTLN--PCVFYKTSQ-- 65  
DB 104 NMCGKMGACNSQGIAYWSTDLFGFTPTNV-TLEMTGYFLPQGTGYFKFATVDOS 162  
QY 66 -----TSVDLTADGVK-----YATCOFYSGEEFTTFSTL 94  
DB 163 AILSVGGATAFNCAQOQPITSTNFTIDGKPMGSGSLPPNIEGTVMYVAGYYPM---- 218  
QY 95 TCTVNDALKSIKAFGT----VTLPFAFNVGGTSGSTDLEDSKCFAGTNTVTNDGDKD 150  
DB 219 -----KVYISNAVSWGTLPISVTLP-----DGTVTSDDFEG-----YVVSFDD---- 256  
QY 151 ISIDVEFEKSTV-DPSAYLYASRVMPSLNKLVTFLVAPQENGYSCTGTMGSSNGDVAI 209  
DB 257 ----DLQSQNCVPPDSNYA-VSTTTTTTTPWTGTTSTSTMTTGTGNGVPTDETIVVI 312  
QY 210 DCSNIHIGITKGLNDWNVSVSEFSYKTCSTNGIOIKYQNPAGYRPFIDAYISATDV 269  
DB 313 RPTTASTIITTEPNNSFTSTSTELTIVTCTNGVTDETII-----VIRPTTATTA 366  
QY 270 NOYTLAYNDYTCAGRSQKPFLLRWGTGKNSDAGSNGI-----VIVATRTVDTSTAVT 326  
DB 367 IITTEPNNSFTSTSTEL-----ITVTGNGLTPTDETIIIVIRPTTATTA 413  
QY 327 TL-BENPSVDKTKT-----TEILOPIPTTIT-----SYGVYIS 361  
DB 414 TTOPNDFTSTSTMTTGTGNGLTPTDETIIVIRPTTATTA 473  
QY 362 YLTAKTA-----PIGETATVVDVPHHTT--TVTSEWGTIT-----TTTTRTN--PTDS 407  
DB 474 MTVTGTNGLTPTDET-IIIVIRPTTATTA 531  
QY 408 IDTVVW 413  
:|:|:|

Db 532 -ETIIV 536  
RESULT 9  
YQ3\_CAEEL  
ID YQ3\_CAEEL STANDARD; PRT: 1251 AA.  
AC Q09550;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE Hypothetical 133.5 kDa protein F26C11.3 in chromosome II.  
GN F26C11.3  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Pelodierinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RA Matthews P.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC -----  
DB EMBL; Z47072; CAAB7369.1; -  
DR WormPep; F26C11.3; CE01561.  
DR InterPro; IPR000436; Sushi\_SCR\_CCP.  
DR Pfam; PF00084; sushi; 1.  
DR SMART; SM00032; CCP; 1.  
KW Hypothetical protein.  
FT DOMAIN 120 414  
SQ SEQUENCE 1251 AA; 133498 MW; 2B959ECA03B9954A CRC64;  
Query Match 7.88; Score 171; DB 1; Length 1251;  
Best Local Similarity 23.88; Pred. No. 0.00041;  
Matches 107; Conservative 60; Mismatches 154; Indels 128; Gaps 23;  
QY 27 PGYPTWNAVGLMSLDGTS-----ANPGDFTLNMPVCFYKTSQTSVDLT 71  
DB 609 PGTITYN---WPTGTTTMLPSGEIILSESIAYPNCT-TVLMQLIYTPSTNKTRETT 663  
QY 72 AD--GVKATCFYSGEEFT----TFSTLTCTVN-----DALKSSIKAF- 109  
DB 664 TDEGCKKSTISSSSKSFSTPTTPSPSGTITYNWPTGTTTTLPSGEIILSESIAFQ 723  
QY 110 --GVVTLPTAFNVGSGSTDLEDSKCFAGTNTVTNDGDKDISIDVEFEKSTVDPSPAY 167  
DB 724 NCTTVLMQLIYN-----PSIN-----KTRTEITDAEGCKKTSSTKISTTPTST- 769  
QY 168 LYASRVMPSLNKLVTFLVAPQENGYSCTGTMGSSNGDVADCSNIHIGITKGLNDWNY 227  
DB 770 --SSKPTPTSTSMWTYNNP--TGCTTTL-----PSGEI----- 800  
QY 228 PVSESESYKTCSTNGIOIKYQNPAGYRPFIDAYISATDVN--OYTLAYNDYTCAGS 285  
DB 801 -ILSESIAKNCCTVLMQLIYN--PSTNKTRE-----TTDAQCGKATSSLSKPTSPSS 854  
QY 286 RLQSKP-FTLRW-TGYKNSDAGSNGIIVVATRTVDTSTAVTTLFPNPSVDKTKT---- 339  
DB 855 STASPPPTTYNPTGTTTTLPSGEIILSESIAKNCCTVLMQLIYNPSTNKTRETTT 914  
QY 340 -----TEILOPIPTTITTSYGVYTSYLTAKTAPICE-----TATVIVDVPY 381  
DB 915 DAQCKKATITPTP---ITTY-NWPTGTTTTLPSGEIILSESIAKNCCTVLMQLIY 970  
QY 382 H-----TTTTVTSEWGTITTTTTTRNPT 405

FT	DOMAIN	353	394	P-TYPE 3.
FT	DOMAIN	402	522	12 X APPROXIMATE TANDEM REPEATS,
FT				THR-RICH.
FT	REPEAT	402	411	3-1.
FT	REPEAT	412	419	3-2.
FT	REPEAT	420	431	3-3.
FT	REPEAT	432	443	3-4.
FT	REPEAT	444	453	3-5.
FT	REPEAT	454	460	3-6.
FT	REPEAT	461	472	3-7.
FT	REPEAT	473	479	3-8.
FT	REPEAT	480	491	3-9.
FT	REPEAT	492	498	3-10.
FT	REPEAT	499	515	3-11.
FT	REPEAT	516	522	3-12.
FT	DOMAIN	525	566	P-TYPE 4.
FT	DOMAIN	572	613	P-TYPE 5.
FT	DOMAIN	620	661	P-TYPE 6.
FT	DISULFID	162	188	BY SIMILARITY.
FT	DISULFID	172	187	BY SIMILARITY.
FT	DISULFID	182	199	BY SIMILARITY.
FT	DISULFID	307	333	BY SIMILARITY.
FT	DISULFID	317	344	BY SIMILARITY.
FT	DISULFID	327	344	BY SIMILARITY.
FT	DISULFID	334	360	BY SIMILARITY.
FT	DISULFID	364	379	BY SIMILARITY.
FT	DISULFID	374	391	BY SIMILARITY.
FT	DISULFID	526	552	BY SIMILARITY.
FT	DISULFID	536	551	BY SIMILARITY.
FT	DISULFID	546	563	BY SIMILARITY.
FT	DISULFID	573	599	BY SIMILARITY.
FT	DISULFID	583	598	BY SIMILARITY.
FT	DISULFID	593	610	BY SIMILARITY.
FT	DISULFID	621	647	BY SIMILARITY.
FT	DISULFID	631	646	BY SIMILARITY.
FT	DISULFID	641	658	BY SIMILARITY.
FT	VARIANT	276	276	K -> E.
FT	VARIANT	354	354	C -> R.
FT	VARIANT	415	415	T -> A.
FT	SEQUENCE	662 AA:	67774 MW:	F085277F1ED2FD40 CRC64:

  

Query Match	7.6%;	Score	167.5;	DB 1;	Length	662;
Best Local Similarity	24.3%;	Pred. NO.	0.00033;			
Matches	94;	Conservative	28;	Mismatches	166;	Gaps

  

QY	42	GTSANPGDTFLNMPCKVKYITSGTSVDLTADGVKATPCFYSGEFTFTEFLCTVND	101
DB	206	GHSHEEHTTTTAKPTTIOATTTTPTTTT---ATPTTTTAKPTT	254
QY	102	LKSSIKAFGVTLPAPNVGSGSSDLEDSCKFTAGTNVTFDGDKDISIDVEFEKST	161
DB	255	TTTTTKATTTTTP-----TTTTTTTKATTTPTT-----TT	286
QY	162	VDPSAYLIASRWPSLNKVTTLFVAPQCEGYTSGTCMFSSNGDVAIDCSNIHIGITKG	221
DB	287	TTPTT-----TTTKATTTT-----TSGECKMPSKRE---DCG--YSGITES	325
QY	222	LDNWNYPVSESFSFKTCSNGIQI-----KYQNPAGYRFDIDAYISATDVNQYTLA	275
DB	326	QCRTKGCCFDSISIPQTKWCFVTLQSVADCKPEPSQRVDCGFRG-----ITADQCRKNCC	380
QY	276	YNDYTCAGSRLOSKPFRLRWGTGYKNSDAGSNGIVIVATRTVTDSTTAVTLPFPNSVD	335
DB	381	F--DSSTSGT-----KCFYSTSQ-----VAATKTTTPTT--TTPTTTTTT	419
QY	336	K---KXTBILQIPTTTTITSVGVTTSYLTKTAPIGETATVIVDVPYHTTIVTISEW	391
DB	420	KATTTPTTTTPTTTTPTTTT-----TKATTTTPTTTPTTTTKATTTTPTTTTPTTTT	478
QY	392	TGTTT--TTTRTNPTDSIDTVVQVP	416
DB	479	TKATTTPTTTTPTTTTPTTTT-----TKATTTPTTTTPTTTT	505



```
RESULT 11
FLO5_YEAST STANDARD: PRT: 1075 AA.
ID AC F38894;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Flocculation protein FLO5 precursor (Flocculin 5).
GN FLO5 OR YHR211W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S488C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hallier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
RN [2]
RP REVIEW
RX MEDLINE=96076625; PubMed=7502576;
RA Teunissen A.W., Steensma H.J.;
RT "Review: the dominant flocculation genes of Saccharomyces cerevisiae
RT constitute a new subtelomeric gene family.";
RL Yeast 11:1001-1013(1995).
CC -|- FUNCTION: MAY BE DIRECTLY INVOLVED IN THE FLOCCULATION PROCESS.
CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC -|- (Potential).
CC -|- PTM: EXTENSIVELY O-GLYCOSYLATED (PROBABLE).
CC -|- SIMILARITY: BELONGS TO THE FLOCCULIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U00029; AAB69731.1;
CC PIR; S48992; S48992.
CC SCD; S0001254; FLO5.
CC InterPro; IPR001389; Flocculin.
CC Pfam; PF00624; Flocculin; 8.
CC Glycoprotein; Membrane; Repeat; Cell wall; Signal; GPI-anchor.
CC SIGNAL 1 24
CC CHAIN 25 ?
CC PROPEP 25 ?
CC CARBOHYD 135 ?
CC CARBOHYD 187 ?
CC CARBOHYD 203 ?
CC CARBOHYD 262 ?
CC CARBOHYD 663 ?
CC CARBOHYD 749 ?
CC SEQUENCE 1075 AA; 111981 MW; D151B370B60C8D9F CRC64;
CC
CC Query Match 7.48; Score 161.5; DB 1; Length 1075;
CC Best Local Similarity 22.28; Pred. No. 0.0015;
CC Matches 110; Conservative 57; Mismatches 209; Indels 119; Gaps 22;
CC
CC 21 NYAFKPG-YPTNVAVLGWSLD-GTSANPGDTFTLNM-----PCVFKYTSQTSDVLT 71
CC
```

```
Db 104 NMKGKMGACNSQGIAYWSTDLFGFYTPTNV-TLEMTGYFLPQTGSGTFSFATVDS 162
QY 72 A-----DGVKATQCFYSGEEFTTSTLTCTVN----- 99
Db 163 AILSVGSGIAFECC-----AQEQPPIITSNFTINGIKPWGDSLPDNIITGVIMYAGVYPL 218
QY 100 DALKSSIKAFGVTVLPPIAFNVGGTSGST-----DLEDSKC-----FTAGTNTV 142
Db 219 KVVYNAVSWGTLPLISVELPDGTVTSDNFEQVYVYFDDLSQSNCTIPDPSIHHTTSTITT 278
QY 143 -----TNDGDKDISIDVEFEKSTVDPSAYL-----ASRVMPSLNKVTTTLFVAPOCE 190
Db 279 TTEPWGTGTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 338
QY 191 NGYTSGTMGFSSSSNGDAIDCSNIHIGITKGLNDMNPVSSSEFSYTKTCTSGNGIQIKYO 250
Db 339 MTTVTGNGQPTDETVIVIRTPTSEGLITTTTTEPWGTGTTSTSTSTSTSTSTSTSTSTSTSTST 397
QY 251 NVPAGYRPFIDAYISATDVNQVTLAYTNDYTCAGSLQSKPFTLRWTKYKNSD----- 303
Db 398 TVIVIRTPTSEGLITTT-----TEPWGTGTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 448
QY 304 -AGSNGIVIVAN---TETVTDSTAVTLPNPSPVDKTKTIELQP-----IPTT----- 349
Db 449 TPSEGLITTTTEPWGTGTTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 508
QY 350 --TITSYGVVTSVLKTAIPGETATVI-VDPYHTTHTTSTSTSTSTSTSTSTSTSTSTSTSTSTST 400
Db 509 TGTFTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 568
QY 401 RTN--PTDSIDRVVV 413
Db 569 GTNGQPTD--ETVIV 581
RESULT 12
CHT3_CANAL STANDARD: PRT: 567 AA.
ID CHT3_CANAL
AC P40954;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Chitinase 3 precursor (EC 3.2.1.14).
GN CHT3.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10261;
RX MEDLINE=95223977; PubMed=7708682;
RA McCreath K.J., Specht C.A., Robbins P.W.;
RT "Molecular cloning and characterization of chitinase genes from
RT Candida albicans.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:2544-2548(1995).
CC -|- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -|- SUBCELLULAR LOCATION: Secreted (probable).
CC -|- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U15801; AAA68016.1;
CC HSSP; P23472; 2HVN.
```



QY 320 -----DSTAVVTLP 329  
 Db 359 SPCKAETLPGDNTVFTHGSLGMDALGDGKVLVLSLMDHADMWLDSYPTTSCA 418  
 QY 330 FNPVDKTKTEIILQIPITPTTITSYV--GVTTSYLT-----KTAPIGET----- 372  
 Db 419 SPSPGVAR-----GTCPTTGNATVEANPNSVYTSNIKFTGLNSTYSGTSSGSSS 471  
 QY 373 -ATVIVDPVHTTTVTSEWGTITTTTTTRTNPDTSDTIVVQV 415  
 Db 472 SSTTL-----TTKASTSTTSKTKTTTTTSKTSSTSSSTNVQOL 509

RESULT 14  
 AMYL\_YEAST STANDARD: PRT: 1367 AA.  
 AC P08640; P08068;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Glucoamylase SI/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).  
 GN STAL OR STA2 OR MAL5 OR YIR019C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RA Barrill B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
 RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
 RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,  
 RA Lewis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
 RA Rajandream M.A., Riles E., Rowley N., Skelton J., Smith V.,  
 RA Walsh S.V., Whitehead S.,  
 RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RC SEQUENCE OF 1-242 AND 762-1331 FROM N.A.  
 RA MEDLINE=87194600; PubMed=3106330;  
 RA Ramashita I., Nakamura M., Fukui S.;  
 RA "Gene fusion is a possible mechanism underlying the evolution of  
 STAL.";  
 RL J. Bacteriol. 169:2142-2149(1987).  
 RN [3]  
 RC SEQUENCE OF 1-31 FROM N.A.  
 RA STRAIN=SPX101-1C;  
 RC MEDLINE=89031230; PubMed=3141213;  
 RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;  
 RA "Similar short elements in the 5' regions of the STA2 and SGA genes  
 from Saccharomyces cerevisiae.";  
 RL FEBS Lett. 239:179-184(1988).  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-  
 CC glucose residues successively from non-reducing ends of the chains  
 CC with release of beta-D-glucose.  
 CC -!- SIMILARITY: TO S.POMBE SPBC215.13.  
 CC -!- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed, usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: Z38061; CAA86176.1;  
 DR EMBL: M16164; AAA35014.1;  
 DR EMBL: M16165; AAA35013.1;  
 DR EMBL: X13657; CAA32069.1;  
 DR PIR: B26877; B26877.  
 DR PIR: A26877; A26877.

DR PIR: S48478; S48478.  
 DR SGO; S0001458; MUC1.  
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;  
 KW Signal; Multigene family.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1367  
 FT DOMAIN 210 1367  
 FT SER/THR-RICH.  
 FT CARBOHYD 817 817  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 874 874  
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;  
 Query Match 7.1%; Score 154.5; DB 1; Length 1367;  
 Best Local Similarity 23.6%; Pred. No. 0.006;  
 Matches 115; Conservative 40; Mismatches 184; Indels 149; Gaps 22;  
 QY 15 TWSNAANYAFKGPYPTWNAVLGWSLDGTSANPGDTEFLNMPCKVFKYTSQ--TSVDLTA 72  
 Db 868 TPSSSSNITSSAP-----SSIPFSSTVESFSTGTTVT--PSSSKYPGSQTETSVSSTT 918  
 QY 73 DG--VKYATCQFYSGEEFTFTLCT--VNDALKS-----SIKAFGTVTLPIAFNVGGTG 124  
 Db 919 ETTIVPTKTTTSVTPSTTTTITTTTCTGTSAGETTSKSPKTV-TTIVPTTTTTSVTT 977  
 QY 125 SSTDLDSKCFAGTNT---VTENDGKDIS----- 152  
 Db 978 SSTTTTITTCSTGTSAGETTSKSPKTTTTPVPCSTSPSESTSTSPPTPTTVV 1037  
 QY 153 ----IDVEFEKSTVD-----PSAYLVASRVMPSLNKVT----TLFVAPQCN 191  
 Db 1038 STTVVTEVSTSKPGCEITTFVTKNIPPTTLTIAPTSTVTVTNFTPTTTTTCST 1097  
 QY 192 GYTSGMFSSSGDVAIDNSNHIIGIKGLNDNMPVSESFYKTCSTNSGIQIKYON 251  
 Db 1098 G-----TNSAGETTSKSPKTVTT-----VPCSTGTGETTEAT- 1133  
 QY 252 VPAGYRPFDAISATDVNOYTLAYND--VTCAGSLQSKPFTLWGTGYNKSDAGSNGI 309  
 Db 1134 -----LVTTAVTTTIVTTESSTGTSAGKTTGYTTSVP-----TIVVTLAPSAVP 1181  
 QY 310 V-----IVATRTVDTSTAVTTLFPNPSVDKTKTI-----EILQPIPTT 349  
 Db 1182 TPATNAVPTTITTECSAATNAAGETTSVCSAKTIIVSSASAGENTAPSAPTPTTAIPT 1241  
 QY 350 TITT-SYVG-----VTTSYLTKTAPIGETATVIDVPYHTTTTIVTSEWGTITTTTTRT 402  
 Db 1242 VITTESVGTNSAGETTTGTGTSIP-----TTVITTLIPSGNARKYETVATAT 1291  
 QY 403 NPTDSIDT 410  
 Db 1292 NPI-SIKT 1298  
 RESULT 15  
 AGAL\_YEAST STANDARD: PRT: 725 AA.  
 ID AGAL\_YEAST  
 AC P32323;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE A-agglutinin attachment subunit precursor.  
 GN AGAL OR YNR044W OR N3431  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RA MEDLINE=91304412; PubMed=2072914;  
 RA Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;  
 RA "The AGAL product is involved in cell surface attachment of the  
 RT Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";  
 RL Mol. Cell. Biol. 11:4196-4206(1991).

RN SEQUENCE FROM N.A.  
RA Pohl T.M.;  
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.  
CC -|- SUBUNIT: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ  
CC SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY  
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,  
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR  
CC AGGREGATION DURING MATING.  
CC -|- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO  
CC A CORE SUBUNIT.  
CC -|- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR  
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).  
CC -|- PTM: EXTENSIVELY O-GLYCOSYLATED.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; M60590; AAA34382.1; -;  
DR EMBL; 271659; CAA96325.1; -;  
DR PIR; S17031; S17031.  
DR PIR; A41258; A41258.  
DR SGD; S0005327; AGA1.  
KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;  
KW Phormone response.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 725 A-AGGLUTININ ATTACHMENT SUBUNIT.  
FT DOMAIN 53 493 2 X APPROXIMATE REPEATS.  
FT REPEAT 53 149 1-1.  
FT REPEAT 395 493 1-2.  
FT DOMAIN 182 307 18 X APPROXIMATE TANDEM REPEATS, SER/THR-  
FT REPEAT 182 188 2-1.  
FT REPEAT 189 195 2-2.  
FT REPEAT 196 202 2-3.  
FT REPEAT 203 209 2-4.  
FT REPEAT 210 216 2-5.  
FT REPEAT 217 223 2-6.  
FT REPEAT 224 230 2-7.  
FT REPEAT 231 237 2-8.  
FT REPEAT 238 244 2-9.  
FT REPEAT 245 251 2-10.  
FT REPEAT 252 258 2-11.  
FT REPEAT 259 265 2-12.  
FT REPEAT 266 272 2-13.  
FT REPEAT 273 279 2-14.  
FT REPEAT 280 286 2-15.  
FT REPEAT 287 293 2-16.  
FT REPEAT 294 300 2-17.  
FT REPEAT 301 307 2-18.  
SQ SEQUENCE 725 AA; 73353 MW; 70420C853B0B01F8 CRC64;

Query Match 7.0%; Score 153.5; DB 1; Length 725;  
Best Local Similarity 21.9%; Pred. No. 0.0032;  
Matches 95; Conservative 64; Mismatches 209; Indels 65; Gaps 15;  
QY 1 AKTIGVDFSNLSLWNSAANYAFKGGYPTWNAVGLWSLDGTSANPGDFTFLNMPGVFK 60  
DB 163 ASIISPVSTLSSTSSN-----PTT-----SLSTSTSPSSTST--SPSSTS 204  
QY 61 YTTQTSVOLTADGVKYATCPQYSGEEFTFTLTCTVNDAL-----KSSIKAFGTVILP 115  
DB 205 TSSSSTSTSSSSTSTSSSSTS--TSPSSSTSSSLTSTSSSSTSTSSSSTSTSP 263  
QY 116 IAFNVGGTGSSTDLDSKCFAGTNTVTF-----NDGDKDISIDVEFEK 159  
DB 264 SSTSTSSSSTSTSPSSKSTSPSSSTSTSSYSTSTSPSLTSSSPTLASTSPSSTSTSTFTD 323

QY 160 STVDPFSAVLASRVMPSL-NKVTTLFVAPOCENGYSGTMGFSSNGDVAIDCSNIHIGI 218  
DB 324 STSSILGSSIASSTSVSLYSPSTPVYSPSTSSNNVATPSM--TSSTVETTVSSOSSEYI 381  
QY 219 TKGLNDMKNYPVSSSEFSYTKTCTNSGIOIKYONYPAGYRPFIDAYISATDVNOYTLAYTN 278  
DB 382 TKSSISTTIPFSNSTYFT---TVSGVTWYTT---WCPYSESESTSTLSMHEVTYTD 434  
QY 279 DYTCA-GSRLOSKPFLRWTKYKNSDAGSNGIVIVATRTVTDSTAVTTLFPNPSVDKT 337  
DB 435 AIVCTHESCMPSQTTSLTSSIKMS---TKNVATSVSTVSSVACSTCAETSHSYSSV 491  
QY 338 KTIELQPIPTTTTTSYGVVTSYLTXTAPIGETATVVDVPYHTTTTTSWTSEWTGIIT 397  
DB 492 QTASSSVYQTQTTSTKSWV---SSMTTSDDEFNKHAT---GKYH-----VTSGTSTIST 540  
QY 398 TTRTNPTDSIDT 410  
DB 541 SVSEATSTSSIDS 553

Search completed: October 3, 2002, 16:34:17  
Job time: 250 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 3, 2002, 16:29:47 ; Search time 31.78 Seconds  
(without alignments)  
2264.503 Million cell updates/sec

Title: US-09-715-876-8\_COPY\_17\_432

Perfect score: 2190  
Sequence: 1 AKTTGVDSPFSLTWSNAA.....TTTTRTNPDTSDTIVVQVP 416

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organellae:\*

9: sp\_plant:\*

10: sp\_protist:\*

11: sp\_riccia:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rviro:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1922	87.8	1270	3 Q9Y743	Q9Y743 candida alb
2	1889	86.3	1047	3 Q93865	Q93865 candida alb
3	1764.5	80.6	468	3 Q9URQ0	Q9URQ0 candida alb
4	1605	73.3	1523	3 Q9HFX4	Q9HFX4 candida alb
5	1604	73.2	469	3 Q9URP8	Q9URP8 candida alb
6	1602.5	73.2	468	3 Q9C471	Q9C471 candida alb
7	1549	70.7	1443	3 Q9H8F2	Q9H8F2 candida alb
8	1471	67.2	336	3 Q9H8F2	Q9H8F2 candida alb
9	1191	54.4	338	3 Q9HF70	Q9HF70 candida alb
10	1176	53.7	331	3 Q9HF69	Q9HF69 candida alb
11	997.5	45.5	2297	3 Q9HCK6	Q9HCK6 candida alb
12	990.5	45.2	433	3 Q9H8F1	Q9H8F1 candida alb
13	981	44.8	353	3 Q9HF71	Q9HF71 candida alb
14	359.5	16.4	126	3 Q9HF49	Q9HF49 candida alb
15	341	15.6	127	3 Q9HF50	Q9HF50 candida alb
16	219	10.0	52	3 Q93861	Q93861 candida alb

17	217	9.9	1195	3 Q96WU8	Q96WU8 schizosacch
18	199	9.1	1752	2 Q9AE52	Q9AE52 ruminococcu
19	198.5	9.1	948	3 Q74346	Q74346 schizosacch
20	180	8.2	1498	2 Q9L448	Q9L448 arthrobacte
21	179.5	8.2	956	5 Q00908	Q00908 cryptospori
22	175.5	8.0	267	3 Q9P6S0	Q9P6S0 schizosacch
23	175.5	8.0	1832	5 Q9H503	Q9H503 cryptospori
24	175	8.0	565	3 Q9HDY9	Q9HDY9 schizosacch
25	172	7.9	1217	4 Q9UKW9	Q9UKW9 homo sapien
26	171.5	7.8	456	3 Q42840	Q42840 schizosacch
27	170	7.8	957	4 Q14651	Q14651 homo sapien
28	167	7.6	1220	3 Q9C0Y2	Q9C0Y2 schizosacch
29	165	7.5	513	4 Q43418	Q43418 homo sapien
30	163	7.4	648	5 Q95QX0	Q95QX0 caenorhabdi
31	162.5	7.4	567	17 Q96VT2	Q96VT2 sulfolobus
32	156.5	7.1	770	5 Q20908	Q20908 caenorhabdi
33	156	7.1	851	5 Q17893	Q17893 caenorhabdi
34	155.5	7.1	1893	5 Q9NKC9	Q9NKC9 drosophila
35	155	7.1	597	16 Q9RY11	Q9RY11 delnoccoccu
36	155	7.1	1180	16 Q9CHH4	Q9CHH4 lactococcu
37	154	7.0	1283	3 Q9USQ3	Q9USQ3 schizosacch
38	154	7.0	1283	3 Q9URU4	Q9URU4 schizosacch
39	153.5	7.0	1795	5 Q76894	Q76894 drosophila
40	153.5	7.0	3570	4 Q99552	Q99552 homo sapien
41	153	7.0	338	4 Q43420	Q43420 homo sapien
42	153	7.0	769	5 Q17921	Q17921 caenorhabdi
43	151	6.9	961	3 Q92223	Q92223 emericella
44	151	6.9	2586	5 Q9VTK8	Q9VTK8 drosophila
45	149	6.8	966	5 Q9U3P1	Q9U3P1 caenorhabdi

## ALIGNMENTS

RESULT	1
Q9Y743	PRELIMINARY; PRT: 1270 AA.
AC	Q9Y743; 1999 (Tremblrel. 12, Created)
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT	01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE	AGGLUTININ-LIKE PROTEIN.
GN	ALS5.
OS	Candida albicans (Yeast).
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX	NCBI_TaxID=5476;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=1161;
RX	MEDLINE=21064501; PubMed=11124701;
RA	Hoyer L.L., Hecht J.E.;
RT	"The ALS5 gene of Candida albicans and analysis of the Als5p N-
RT	terminal domain.";
RL	Yeast 18:49-60(2001).
DR	EMBL; AF068866; AAD32849.1; ..
SQ	SEQUENCE 1270 AA; 133337 MW; D8E5FA5853F6D5C5 CRC64;

Query Match 87.8%; Score 1922; DB 3; Length 1270;  
Best Local Similarity 87.0%; Pred. No. 1.7e-116;  
Matches 362; Conservative 20; Mismatches 34; Indels 0; Gaps 0;

QY	1	AKTTGVDSPFSLTWSNAAAFKPGYPTWNAVGLWSLDGTSANBCDFTTLNMCVFK	60
Db	17	AKATGFGNSDLSLWSNAAAFKPGYPTWNAVGLWSLDGTSANBCDFTTLNMCVFK	76
QY	61	YTSQTSVDLTADGVKKVATCFVSGEEFTTSTCTVNDALKSSIKAFGTVTLPIAFNV	120
Db	77	FTASQSVDLTADGVKKVATCFVSGEEFTTSTCTVNDALKSSIKAFGTVTLPIAFNV	136
QY	121	GGTGSSTDLDSKCFAGTNTVTFNDGDKDISIDVEKSTVDPSSAYLYASRYMPSLNKY	180

```
Db 137 GGTGSTDLEDSKCFAGTAGINTVTFNDGSKLSIAVNFPEKSTVDRSGLTTSREMPSLNKI 196
Qy 181 TTLFVAPQCENGYTSMTGFFSSNGDVAIDCSNIHIGITKGLNDWNPVSESFSTKTC 240
Db 197 ATLYVAPQCENGYTSMTGFFSNGDVAIDCSNHVIGISKGVNDWNPVSESFSTKSC 256
Qy 241 TSNQIQIKYQNVAGYRPFIDAYISATDVNQYTLATYNDYTCAGSRQSKPFTLRWTGYK 300
Db 257 SSFGISITYQNVAGYRPFIDAYISPDNNQYQLSKYNDYTCDDYQWCHAPFTLRWTGYK 316
Qy 301 NSDAGSNGIVATRTVTDSTAVTTLPPNPVSKTKTIELQPIPTTTTTSYGVGVT 360
Db 317 NSDAGSNGIVATRTVTDSTAVTTLPPNPVSKTKTIELQPIPTTTTTSYGVGVT 376
Qy 361 SYLTAPIGETATVIVDPYHTTTVTSEWTKTITTTTTRNPTDSIDTVVQVP 416
Db 377 SYSTKAPIGETATVIVDPYHTTTVTSEWTKTITTTTTRNPTDSIDTVVQVP 432

RESULT 2
Q93865
ID O93865 PRELIMINARY; PRT: 1047 AA.
AC O93865;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AGGLUTININ-LIKE CELL SURFACE PROTEIN.
GN ALS8.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC10261;
RA Leng P., Lee P.R., Wishart J.A., Wu H., Brown A.J.P.;
RT "Sequence of the hypha-specific, agglutinin-like cell surface protein,
RT ALS8 from Candida albicans."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF051313; AAD02580.1; -;
SQ SEQUENCE 1047 AA; 111944 MW; C2327659AA911P2E CRC64;

Query Match 86.3%; Score 1889; DB 3; Length 1047;
Best Local Similarity 84.9%; Pred. No. 1.9e-114;
Matches 353; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

1 AKTITGVDFSNLTSNAANYAFKPGYPTWNAVGLWSLDGTSANPGDTFTLNMPGVK 60
Db 17 AKTITGVDFSNLTSNAANYAFKPGYPTWNAVGLWSLDGTSANPGDTFTLNMPGVK 76
Qy 61 YTTSTQSVDLTADGVKYATCQYSGEETFTSTLTCTVNDALKSSIKAFGTVTLPIAFNV 120
Db 77 FTTSTQSVDLTADGVKYATCQYSGEETFTSTLTCTVNDALKSGTGTVTLPIAFNV 136
Qy 121 GGTGSTDLEDSKCFAGTAGINTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 180
Db 137 GGTGSTDLEDSKCFAGTAGINTVTFNDGDKDISIDVEFEKSTVDPSAYLYASRVMPSLNKV 196
Qy 181 TTLFVAPQCENGYTSMTGFFSSNGDVAIDCSNIHIGITKGLNDWNPVSESFSTKTC 240
Db 197 ATLYVAPQCENGYTSMTGFFSNGDVAIDCSNHVIGISKGVNDWNPVSESFSTKSC 256
Qy 241 TSNQIQIKYQNVAGYRPFIDAYISATDVNQYTLATYNDYTCAGSRQSKPFTLRWTGYK 300
Db 257 SSFGISITYQNVAGYRPFIDAYISPDNNQYQLSKYNDYTCDDYQWCHAPFTLRWTGYK 316
Qy 301 NSDAGSNGIVATRTVTDSTAVTTLPPNPVSKTKTIELQPIPTTTTTSYGVGVT 360
Db 317 NSDAGSNGIVATRTVTDSTAVTTLPPNPVSKTKTIELQPIPTTTTTSYGVGVT 376
Qy 361 SYLTAPIGETATVIVDPYHTTTVTSEWTKTITTTTTRNPTDSIDTVVQVP 416
Db 377 SYSTKAPIGETATVIVDPYHTTTVTSEWTKTITTTTTRNPTDSIDTVVQVP 432

RESULT 4
Q9HEX4
ID Q9HEX4 PRELIMINARY; PRT: 1523 AA.
AC Q9HEX4;
DT 01-NAR-2001 (TrEMBLrel. 16, Created)
DT 01-NAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-NAR-2001 (TrEMBLrel. 16, Last annotation update)
DE AGGLUTININ-LIKE PROTEIN (FRAGMENT).
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
```



OX		NCBI_TaxID=5476;	
RN	[1]		
RP		SEQUENCE FROM N.A.	
RC		STRAIN-SCS314;	
RA		Chen X., Chen J.-Y.;	
RL		*ALS4 (agglutinin-like sequence) of Candida albicans.*;	
RT		submitted (MAY-2000) to the EMBL/GenBank/JDBJ databases.	
RF		EMBL: AF272027; AAG25054.1; -	
FT		NON_TER 1523 1523	
SQ		SEQUENCE 1523 AA; 159168 MW; 73AF3B3E442FD53C CRC64;	
		Query Match 73.3%; Score 1605; DB 3; Length 1523;	
		Best Local Similarity 70.9%; Pred. No. 7.3e-96;	
		Matches 295; Conservative 45; Mismatches 76; Indels 0; Gaps 0;	
Oy	1	AKTITGVDFSEFLWSNAANFAKPGCYPTWNAVGLWSLDGTSANPGDTFTLNMPVFK 60	
Db	17	AKVITIGFDSFLSWTNAASYSRGPNPMTAVIGSLDGTASAGDTFTLDMPCVK 76	
Oy	61	YTTSQTSDLTADGVKATCOFYSGEEFTFTSLCTVNDAKASSIKAFGTGTLPTAEV 120	
Db	77	FITDQTSIDLVDGRVTATNLNSAEEFTTFSSVCTVTTMTADTKAIGTVTLTPFSFV 136	
Oy	121	GGTGSGTDLESKCFCTAGTNVTTFNDGDKDISIDEVEKSTVDPSAYILASRYMSLANV 180	
Db	137	GGSGSDVDLANSQCFTAGINTVFNDGDTISATVDFEKSTVAASDRLLSRLLPSLSQA 196	
Oy	181	TTLFVAPOCENGYTCSTMFGSSGDVAIDCSNIHGIKTGLANDNYPVSESFYKTC 240	
Db	197	VSLFVPOECANGYTCGMFGSTAGTCATIDCSIVHGINSGLANDNYPISESEFTIYIC 256	
Oy	241	TSNGIQKYONWPAGYRPFIDAYISATDVNQYTLAYNTDTCAGSLQSPFFTLRWGYK 300	
Db	257	TSTSVLYTONWPAGYRPFVDAYSATSRSYAMRYTNITACVGAAVSDDSEFTHWWSGS 316	
Oy	301	NSDAGSGNVIATRVTVDSSTAVTLPLFPNPVSOKTKTEILOPIPTTTITSYVGVT 360	
Db	317	NSQAGSNGITVWTRVTVDSSTAVTLPLFPNPSESOKTKTEILOPIPTTTITSYVGVT 376	
Oy	361	SYLTAKPIGETATVINDVPYHTTPTVSEWTCITTTTTNTPTDSDIAVVVQVP 416	
Db	377	SVSTKTAPIGETATVINDVPYHTTPTVSEWTCITTTTTNTPTDSDIAVVVQVP 432	
RESULT	5		
ID	QURPB8	PRELIMINARY; PRT: 469 AA.	
QU	QURPB8		
AC	QURPB8;		
CD	01-MAY-2000 (TREMBLrel. 13, Created)		
DE	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAY-2000 (TREMBLrel. 13, Last annotation update)		
GN	AGGLUTININ-LIKE PROTEIN (FRAGMENT).		
DN	ALS4.		
GC	Candida albicans (Yeasts).		
OS	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
OC	Saccharomycetales; mitosporic Saccharomycetales; Candida.		
OX	NCBI_TaxID=5476;		
RN	[1]		
RP		SEQUENCE FROM N.A.	
RC		STRAIN=1161;	
RA		MEDLINE=98440424; PubMed=9765564;	
RX	Hoyer J.L., Payne T.I., Recht J.E.;		
RT	"Identification of candida albicans ALS2 and ALS4 and localization of		
RR	als proteins to the fungal cell surface.";		
RL	J. Bacteriol. 180:5334-5343(1998).		
RD	EMBL: AF024586; AAC64241.1; -		
FT	NON_TER 469 469		
SQ	SEQUENCE 469 AA; 49597 MW; 88BC96D79142C8DB CRC64;		
	Query Match 73.2%; Score 1604; DB 3; Length 469;		
	Best Local Similarity 71.2%; Pred. No. 2.2e-96;		

Matches	296;	Conservative	43;	Mismatches	77;	Indels	0;	Gaps	0;														
Qy	1	AKTITGVDFSNLTSNAANYAFKPGCYPTNNVNLGWSLDGTSANPGDFTFLNMP	CVK	60																			
Db	17	AKVITGVFNSLTWNAANASPYRGATPTTFAVIGMSLDGATASAGDTFLDMP	CVK	76																			
Qy	61	YTTSTQSVDLTADGVKATQCYQSGEEPTFTSLCTVNDALAKSSIKAFGTVTLP	IAFN	120																			
Db	77	FITDQTSIDLVDAGRTATCNLSAEBEFTFSSVSGTITTTMTADTKAMGTVTLP	FSFV	136																			
Qy	121	GGTGSSTDLDSKCFKTAGTNTVTFNDGDKOISIDVFEKSTVDPSPSAYLYAS	RVMP	180																			
Db	137	GGSGSDVDLANSQCFTAGTNTVTFNDGDTISITVDSEKSTVASSDRILLR	ILPSQA	196																			
Qy	181	TTLFLVAPQECNYSCTGTMGFSSNGDVAIDCSNIHIGITKGLNDWNPVSS	ESFYTK	240																			
Db	197	VSLFLPQECANGYSGTMGFSTAGTGATIDCSVHVGLSGLNDWNPVSS	ESFYTK	256																			
Qy	241	TSNGIOIKYQNVNYPAGYRPFIDAYISATVDVNOYTLAVTNDVYTCAGS	RLOK	300																			
Db	257	TSTSLVLTQNVNYPAGYRPFVDAYVSATRVSSYTMQYTNVYACVGAAS	VDSFT	316																			
Qy	301	NSDAGSGLIVATVTRVTDSTATTVTLPNPNSVOKTKTIELLOPPTTTIT	TSYVGV	360																			
Db	317	NSQAGSNGITIVTTRVTDSTATTVTLPNSDTKTKTIELLOPPTTTIT	SYVGV	376																			
Qy	361	SYLTKTAPIGENATVIVDPYKHTTTVTTSKWTCTITTTTTPDTSIDTV	VQVP	416																			
Db	377	SYSTKAPIGETATVIVDPYHTTTVTTSWTGTTTTTTTNTDTSIDTV	VQVP	432																			
RESULT 6																							
ID	Q9C471	PRELIMINARY;		PRF;		468 AA.																	
AC	Q9C471;																						
DT	01-JUN-2001	(TremBurel. 17, Created)																					
DT	01-JUN-2001	(TremBurel. 17, Last sequence update)																					
DT	01-JUN-2001	(TremBurel. 17, Last annotation update)																					
DE	AGGLUTININ-LIKE	PROTEIN (FRAGMENT).																					
GN	ALS9.																						
OS	Candida albicans (Yeast).																						
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;																						
OX	Saccharomycetales; mitosporic Saccharomycetales; Candida.																						
NCBI_TaxID	5476;																						
[1]																							
SEQUENCE	FROM N.A.																						
RA	STRAIN=1161;																						
RC	Hoyer L.L., Hecht J.E., Mirus K.A.;																						
RL	"The ALS9 gene of Candida albicans.";																						
RT	Submitted (Jan-2000) to the EMBL/GenBank/DBJ databases.																						
DR	EMBL; AF229899; AAK00764.1; -.																						
FT	NON_TER 468																						
FT	SEQUENCE 468 AA; 50127 MW; B291D3EB15F896DE CRC64;																						
Query Match 73.2%; Score 1602.5; DB 3; Length 468;																							
Best Local Similarity 71.6%; Pred. No. 2.7e-96;																							
Matches	298;	Conservative	48;	Mismatches	69;	Indels	1;	Gaps	1;														
Qy	1	AKTITGVDFSNLTSNAANYAFKPGCYPTNNVNLGWSLDGTSANPGDFTFLNMP	CVK	60																			
Db	17	AKVITGVFNSDLTWRVETAYKGPPTNNVNLGWSLMSLTADGDTFLILPC	VK	76																			
Qy	61	YTTSTQSVDLTADGVKATQCYQSGEEPTFTSLCTVNDALAKSSIKAFGTVTLP	IAFN	120																			
Db	77	FITQTSIDLVDAGVATCDNAGEEPTFSSLSCTVNSVSVYARVSGTVKLP																					



RT "Evidence suggesting the presence of an ALS gene family in Candida  
RT dubliniensis and Candida tropicalis.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF201685; AAG35603.1; -;  
FT NON\_TER 1  
FT SEQUENCE 338 338  
SQ SEQUENCE 338 AA; 35908 MW; C0A7F2F94609E172 CRC64;

Query Match 54.4%; Score 1191; DB 3; Length 338;  
Best Local Similarity 64.3%; Pred. No. 8e-70;  
Matches 219; Conservative 51; Mismatches 66; Indels 2; Gaps 1;

QY 45 ANPGDTFTLWPCVFKYTTTSOTSDLTADGVKATCFYSGEEFTFTSLCTVNDALAS 104  
DB 1 ANAGDTFTLWPCVFKYTTTSOTSDLTADGVKATCFYSGEEFTFTSLCTVNDALAS 104  
QY 105 STKAFGTVTLPIAFNPGVGTGSDTLEDSKCTAGTNTVTNDCDKDISIDVEPEKSTVDP 164  
DB 61 NIKAFGVIRPISFNNGTGSVNLQDSKCFRTAGTNSVTIDGDKISIPVDFPKPSS 120  
QY 165 SAYLYASRVMSLUNKVTLFVAPOCENGTSCTMGFSSNGDVAIDCSNIHIGITKGLND 224  
DB 121 SGLIKSRVPTLKLSSLAVASQTAGYKSGVGFSAATNDVTIECSNVHVGITGLNS 180  
QY 225 WNPVSESFYTKTCTSGNGIOIKYQNPAGYRPFIDAYISAPDVNQ--YTLAVTNDYTC 282  
DB 181 WNPVSDSFYTKTCTSSFIITENYENPAGYRPFIDYVYKTKSTTSGFNLVNTSYVC 240  
QY 283 AGSRLQSKPFTLRWGTGYKNSDAGSNGIVATRTVTDSTAVTTLFPNPSVDKTKTIEI 342  
DB 241 TDGKKGNDPLIYFNTSYNSDAGSDGAVIVTKVTVDSTAITLTPDPVTKTIEV 300  
QY 343 LQIPITTTITTSYGVVTSYLTKTAPIGETATVIVDVP 380  
DB 301 LEPIPTTTITTSYGVVTSYLTKTAPIGETATVIVDVP 338

RESULT 10

Q9HF69  
ID Q9HF69 PRELIMINARY; PRT; 331 AA.  
AC Q9HF69;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE AGGLUTININ-LIKE PROTEIN ALSD3P (FRAGMENT).  
GN ALS3.  
OS Candida dubliniensis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=42374;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CD36;  
RA Hoyer L.L., Hecht J.E., En J., Kapteyn J.C., Kils F.M.;  
RT "Evidence suggesting the presence of an ALS gene family in Candida  
RT dubliniensis and Candida tropicalis.";  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF202530; AAG35624.2; -;  
FT NON\_TER 1  
FT SEQUENCE 331 331  
SQ SEQUENCE 331 AA; 35297 MW; 469AB72F9CE029BF CRC64;

Query Match 53.7%; Score 1176; DB 3; Length 331;  
Best Local Similarity 64.3%; Pred. No. 8e-69;  
Matches 214; Conservative 44; Mismatches 73; Indels 2; Gaps 1;

QY 45 ANPGDTFTLWPCVFKYTTTSOTSDLTADGVKATCFYSGEEFTFTSLCTVNDALAS 104  
DB 1 ANAGDTFTLWPCVFKYTTTSOTSDLTADGVKATCFYSGEEFTFTSLCTVNDALAS 104  
QY 105 STKAFGTVTLPIAFNPGVGTGSDTLEDSKCTAGTNTVTNDCDKDISIDVEPEKSTVDP 164

DB 61 QTRALGTVTLPLSEFNIIGSGSDVDITSSQCFKECTNTVTFNDGDTTSTANFORSDVNA 120  
QY 165 SAYLYASRVMSLUNKVTLFVAPOCENGTSCTMGFSSNGDVAIDCSNIHIGITKGLND 224  
DB 121 NDRILLRSLPLSLAKSVTIFIPRCASGYSSGTMGFTAGTDAIIDCSTVHAGISNGLND 180  
QY 225 WNPVSESFYTKTCTSGNGIOIKYQNPAGYRPFIDAYISATDNOYTLAYTNDYTCAG 284  
DB 181 WNPVSESFYTKTCTSGNGIOIKYQNPAGYRPFIDAYISA--LTSYTMQYNTQYTCVG 238  
QY 285 SRLOSKPFTLRWGTGYKNSDAGSNGIVATRTVTDSTAVTTLFPNPSVDKTKTIEI 344  
DB 239 ARPVDASFYNLGYDNAEAGSRGITVVTSTVTDSTAVTTLFPNPSVDVDRKTIIVLQ 298  
QY 345 PIPTTTITTSYGVVTSYLTKTAPIGETATVIV 377  
DB 299 PIPTTTITTSYGVVTSYLTKTAPIGETATVIV 331  
RESULT 11  
Q9HGK6  
ID Q9HGK6 PRELIMINARY; PRT; 2297 AA.  
AC Q9HGK6;  
DT 01-MAR-2001 (TREMBlrel. 16, Created)  
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE AGGLUTININ-LIKE PROTEIN ALS7P.  
GN ALS7.  
OS Candida albicans (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5476;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1161;  
RA Hoyer L.L., Hecht J.E.;  
RT "The ALS6 and ALS7 genes of Candida albicans.";  
RL Yeast 16:847-855(2000).  
DR EMBL; AF201684; AAF98068.1; -;  
SQ SEQUENCE 2297 AA; 244723 MW; 59B020C63027F651 CRC64;

Query Match 45.5%; Score 997.5; DB 3; Length 2297;  
Best Local Similarity 47.0%; Pred. No. 2.8e-56;  
Matches 196; Conservative 71; Mismatches 149; Indels 1; Gaps 1;  
QY 1 AKTTGTFDSFNSLTFWNAANAFKGGPYTWNNAVIGNSLDGTSANPGDFTTLNMPGVFK 60  
DB 18 SKEVTGVNFQNFNSLWSTYRAREEISTLTANAQLEWALDGTIASPGDFTTLVMPGVFK 77  
QY 61 YTTSTQSDLTADGVKATCFYSGEEFTFTSLCTVNDALKSSIKAFGTVTLPIAFNV 120  
DB 78 FMTYTSVOLNTANSTATCFDAGBDTKSFSLKCTVTDLTEDTSVFGSVILPIAFNV 137  
QY 121 GGTGSDTLEDSKCTAGTNTVTFNDGDKDISIDVEPEKSTVDPVSAYLYASRVMSLANK 180  
DB 138 GGSCKSSTTDSKCFSGYNTVTFDGNQLSTTANFLPRRELAFGLVVSQRLMSLDTM 197  
QY 181 TTFVAPOCENGTSCTMGFSSNGDVAIDCSNIHIGITKGLNDWNPVSESFYTKTC 240  
DB 198 TNFVMTPCFMYQSGKLGFTSNDDEIDCDSIHVGITNEINDWMPVSSVPFDHTIRC 257  
QY 241 TSNGIOIKYQNPAGYRPFIDAYISATDNOYTLAYTNDYTCAGSRLQSKPFTLRWT-GY 299  
DB 258 TSRLALIEKTPAGYRPFIDAVIQIPTEPFVKYTNFEEACVNGIYTSIPFTSFESQPI 317  
QY 300 KNSDAGSNGIVATRTVTDSTAVTTLFPNPSVDKTKTIEIQTPTTTITTSYGVV 359  
DB 318 LYDALALAGADLVRTSTVIGSTRTTTLFPFISRLQTKTILVLEPIPTTITTHHGED 377  
QY 360 TSYLTKTAPIGETATVIVDVPYVTTTTTTVTSEWCTGTTTTTTRNPTDSIDTVVQVVP 416



Best Local Similarity 55.1%; Pred. No. 2.1e-16;  
Matches 70; Conservative 15; Mismatches 41; Indels 1; Gaps 1;

Qy 254 AGYRPFDAISATDYNQYTLAYTNDYTCAGSRQSKPFTLRWTGYKNSDAGSNGIVIVA 313  
Db 1 AGYRPFDAALVQAPS-SDYAIQYTKYRCESVQRDDSQKISWAGYTNSDPDSNGAVVVL 59  
Qy 314 TTRVTDSSTAVTTPFPNPSVDKTKIEILOPIPTTTTTSYVGVTTSTYLTAKTAPIGETA 373  
Db 60 TITCTQSTNTIVTLPFPNPTADHKIEIVIPITVTITTSYIGVTTSTYLTATIGDTA 119

Qy 374 TVIVDVP 380  
Db 120 TLVIDMP 126

RESULT 15

Q9HF50 PRELIMINARY; PRT: 127 AA.  
AC Q9HF50;  
AT 01-MAR-2001 (TREMBLrel. 16, Created)  
RX 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE AGGLUTININ-LIKE PROTEIN ALST2P (FRAGMENT).  
GN ALST2.  
OS Candida tropicalis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
OX NCBI\_TaxID=5482;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 13803;  
RX MEDLINE=21186039; PubMed=11290712;  
RA Hoyer L.L., Fundyga R., Hecht J.E., Kapteyn J.C., Klis F.M.,  
RA Arnold J.;  
RT "Characterization of agglutinin-like sequence genes from non-albicans  
candida and phylogenetic analysis of the ALS family.";  
RL Genetics 157:1355-1367(2001).  
DR EMBL; AF211865; AAG43532.1; -.  
FT NON\_TER 1  
FT NON\_TER 127  
SQ SEQUENCE 127 AA; 13668 MW; 178E25AA919EF387 CRC64;

Query Match 15.6%; Score 341; DB 3; Length 127;  
Best Local Similarity 51.2%; Pred. No. 3.4e-15;  
Matches 65; Conservative 22; Mismatches 40; Indels 0; Gaps 0;

Qy 254 AGYRPFDAISATDYNQYTLAYTNDYTCAGSRQSKPFTLRWTGYKNSDAGSNGIVIVA 313  
Db 1 AGYRPFVDVLFSTASDIFTMLTYNEVGADGVYDASMKKANKSYQDLPSPGDGAIIV 60  
Qy 314 TTRVTDSSTAVTTPFPNPSVDKTKIEILOPIPTTTTTSYVGVTTSTYLTAKTAPIGETA 373  
Db 61 TTRCTQSTAVSTLPYDPEIDLTKIEIVLPIPTTTTTSYLGVSYYSTIATIGDTA 120

Qy 374 TVIVDVP 380  
Db 121 TLVIDMP 127

**THIS PAGE BLANK (USPTO)**